(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 February 2005 (17.02.2005)

PCT

(10) International Publication Number
WO 2005/015158 A2

(51) International Patent Classification[7]: G01N

(21) International Application Number: PCT/US2004/025459

(22) International Filing Date: 6 August 2004 (06.08.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/494,071	6 August 2003 (06.08.2003)	US
60/552,064	9 March 2004 (09.03.2004)	US

(71) Applicant *(for all designated States except US)*: **SENOMYX INC.** [US/US]; 11099 N. Torrey Pines Road, La Jolla, California 92037 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **LI, Xiaodong** [CN/US]; 1664 Seattle Slew Way, Oceanside, CA 92057 (US). **STASZEWSKI, Lena** [SE/US]; 17179 Bernardo Center Drive, San Diego, CA 92128 (US). **XU, Hong** [CN/US]; 3955 Via Holgura, San Diego, CA 92130 (US).

(74) Agents: **CLEVELAND, Janell, T.** et al.; Needle & Rosenberg, P.C., Suite 1000, 999 Peachtree Street, Atlanta, GA 30309-3915 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2005/015158 A2

(54) Title: T1R HETERO-OLIGOMERIC TASTE RECEPTORS, CELL LINES THAT EXPRESS SAID RECEPTORS, AND TASTE COMPOUNDS

(57) **Abstract:** The invention relates to compounds that specifically bind a T1R1/T1R3 or T1R2/T1R3 receptor or fragments or subunits thereof. The present invention also relates to the use of hetero-oligomeric and chimeric taste receptors comprising T1R1/T1R3 and T1R2/T1R3 in assays to identify compounds that respectively respond to umami taste stimuli and sweet taste stimuli. Further, the invention relates to the constitutive of cell lines that stably or transiently co-express a combination of T1R1 and T1R3; or T1R2 and T1R3; under constitutive or inducible conditions. The use of these cells lines in cell-based assays to identify umami and sweet taste modulatory compounds is also provided, particularly high throughput screening assays that detect receptor activity by use of fluorometric imaging.

T1R HETERO-OLIGOMERIC TASTE RECEPTORS, CELL LINES THAT EXPRESS SAID RECEPTORS, AND TASTE COMPOUNDS

Cross Reference to Related Applications

This application claims priority to U.S. Provisional Application Serial No 60/494,071 filed on August 6, 2003, and U.S. Provisional Application Serial No 60/552,064 filed March 9, 2004, both of which are incorporated by reference in their entirety.

Background of the Invention

Field of the Invention

The present invention in part relates to the discovery that the T1R receptors assemble to form functional taste receptors. Particularly, it has been discovered that co-expression of T1R1 and T1R3 results in a taste receptor that responds to umami taste stimuli, including monosodium glutamate. Also, it has been discovered that co-expression of the T1R2 and T1R3 receptors results in a taste receptor that responds to sweet taste stimuli including naturally occurring and artificial sweeteners.

Also, the present invention relates to the use of hetero-oligomeric taste receptors comprising T1R1/T1R3 and T1R2/T1R3 in assays to identify compounds that respectively respond to umami taste stimuli and sweet taste stimuli.

The invention also relates to chimeras and truncated versions of T1R1, T1R2, and T1R3, as well as chimeras of T1R1/T1R3 and T1R2/T1R3 receptors comprising human, rat, or human and rat subunits.

Further, the invention relates to the construction of cell lines that stably or transiently co-express a combination of T1R1 and T1R3; or T1R2 and T1R3, including truncated or chimeric versions of these subunits as well as chimeric receptors comprising wild-type or chimeric subunits; under constitutive or inducible conditions.

The use of these cell lines in cell-based assays to identify umami and sweet taste modulatory compounds is also provided, particularly high throughput screening assays that detect receptor activity by the use of fluorometric imaging.

The invention also relates to compounds that bind to T1R1/T1R3, T1R2/T1R3 receptors, as well as T1R1, T1R2, and T1R3 chimeric and truncated subunits and chimeric receptors.

Description of the Related Art

The taste system provides sensory information about the chemical composition of the external world. Mammals are believed to have at least five basic taste modalities: sweet, bitter, sour, salty, and umami. *See, e.g.*, Kawamura *et al.*, *Introduction to Umami: A Basic Taste* (1987); Kinnamon *et al.*, *Ann. Rev. Physiol.*, 54:715-31 (1992); Lindemann, *Physiol. Rev.*, 76:718-66 (1996); Stewart *et al.*, *Am. J. Physiol.*, 272:1-26(1997). Each taste modality is thought to be mediated by a distinct protein receptor or receptors that are expressed in taste receptor cells found on the surface of the tongue (Lindemann, Physol. Rev. 76:718-716 (1996)). The taste receptors that recognize bitter, sweet, and umami taste stimuli belong to the G-protein-coupled receptor (GPCR) superfamily (Hoon *et al.*, *Cell* 96:451 (1999); Adler *et al.*, *Cell* 100:693 (2000)). (Other taste modalities are believed to be mediated by ion channels.)

G protein-coupled receptors mediate many other physiological functions, such as endocrine function, exocrine function, heart rate, lipolysis, and carbohydrate metabolism. The biochemical analysis and molecular cloning of a number of such receptors has revealed many basic principles regarding the function of these receptors. For example, United States Patent No. 5,691,188 describes how upon a ligand binding to a GPCR, the receptor undergoes a conformational change leading to activation of a heterotrimeric G protein by promoting the displacement of bound GDP by GTP on the surface of the $G\alpha$ subunit and subsequent dissociation of the $G\alpha$ subunit from the $G\beta$ and $G\gamma$ subunits. The free $G\alpha$ subunits and $G\beta\gamma$ complexes activate downstream elements of a variety of signal transduction pathways.

The T1R receptors were previously hypothesized to function as sweet taste receptors (Hoon et al., *Cell* 96:541-51 (1999); Kitagawa et al., *Biochem Biophys Res. Commun.* 283:236-42 (2001); Max et al., *Nat. Genet.* 28:58-63 (2001); Montmayeur et al., *Nat. Neurosci.* 4: 412-8 (2001); Sainz et al., *J. Neurochem.* 77: 896-903 (2001)), and Nelson et al. (2001) and Li et al (2002) have recently demonstrated that rat and human, respectively, T1R2 and T1R3 act in combination to recognize sweet taste stimuli.

However, there remains in the art a need for new and improved flavoring agents. For example, one of the five known basic tastes is the "savory" or "umami"

flavor of monosodium glutamate ("MSG"). MSG is known to produce adverse reactions in some people, but very little progress has been made in identifying artificial substitutes for MSG. It is known that a few naturally occurring materials can increase or enhance the effectiveness of MSG as a savory flavoring agent, so that less MSG would be needed for a given flavoring application. For example the naturally occurring nucleotide compounds inosine monophosphate (IMP) or guanosine monophosphate (GMP) are known to have a multiplier effect on the savory taste of MSG, but IMP and GMP are very difficult and expensive to isolate and purify from natural sources, or synthesize, and hence have only limited practical application to most commercial needs in food or medicinal compositions. Less expensive compounds that would provide the flavor of MSG itself, or enhance the effectiveness of any MSG that is present could be of very high value. Similarly, discovery of compounds that are either new "High Intensity" sweeteners (*i.e.* they are many times sweeter than sucrose) would be of value.

What is needed in the art is the identification and characterization of taste receptors which function as sweet and umami receptors, assays for identifying compounds that modulate (enhance or block) sweet and umami taste, and the compounds that specifically bind to these receptors.

Summary of the Invention

The present invention provides chimeric receptors comprising various combinations of human and rat TIRs, such as a chimeric T1R2/T1R3 receptor comprising a human T1R2 subunit and a rat T1R3 subunit; a chimeric T1R2/T1R3 receptor comprising a rat T1R2 subunit and a human T1R3 subunit; a chimeric T1R2 receptor subunit comprising a human extracellular domain, a rat transmembrane domain and a rat intracellular domain; and a chimeric T1R3 receptor subunit comprising a rat extracellular domain, a human transmembrane domain and a human intracellular domain.

The present invention also provides compounds that specifically bind to T1R1, T1R2, T1R3, T1R1/T1R3 and T1R2/T1R3, or isolated subunits, fragments, chimeras or truncated versions thereof as disclosed herein.

The present invention relates to the discovery that different combinations of T1Rs, when co-expressed, produce functional taste receptors that respond to taste

stimuli. Particularly, the present invention relates to the discovery that co-expression of T1R2 and T1R3 results in a hetero-oligomeric taste receptor that responds to sweet taste stimuli. Also, the present invention relates to the discovery that the co-expression of T1R1 and T1R3 results in a hetero-oligomeric taste receptor that responds to umami taste stimuli such as monosodium glutamate.

The present invention also relates to cell lines that co-express T1R1 and T1R3, including human or rat, or T1R2 and T1R3, including human or rat. In preferred embodiments these cell lines will express elevated amounts of the receptors, either constitutively or inducibly. These cell lines include cells that transiently or stably express T1R1 and T1R3 or T1R2 and T1R3.

Also, the present invention provides assays, preferably high throughput screening assays, that utilize the T1R2/T1R3 taste receptor, or the T1R1/T1R3 receptor, preferably high throughput cell-based assays, to identify compounds that modulate sweet or umami taste. The invention also provides assays that include taste tests to confirm that these compounds modulate sweet or umami taste.

The invention also relates to compounds that bind to the N-terminal extracellular domain of T1R2, compounds that bind to the cysteine-rich domain of T1R2, compounds that bind to the Transmembrane Domain of T1R2, compounds that bind to the Transmembrane Domain of T1R3, compounds that bind to the Transmembrane Domain of T1R2 of a truncated receptor h2TM/h3TM, and compounds that bind to the Transmembrane Domain of T1R3 of a truncated receptor h2TM/h3TM, for example.

Brief Description of the Figures

Figure 1 contains a sequence alignment of human and rat T1Rs, human calcium-sensing receptor and rat metabotropic glutamate receptor.

Figure 2 contains RT-PCR amplification experimental results which show that hT1R2 and hT1R3 are expressed in taste tissue.

Figure 3a – 3b contain functional data (intracellular calcium responses) elicited by different sweet taste stimuli in HEK cells stably expressing $G_{\alpha 15}$ that are transiently transfected with human T1R2, T1R3 and T1R2/T1R3 at various concentrations of sweet taste stimuli (Figure 3a); human T1R2/T1R3 dose responses for several sweet taste stimuli (Figure 3b); human T1R2/T1R3 responses to sucrose in

the presence of gurmarin, and endogenous β2-adrenergic receptor responses to isoproterenol in the presence of gurmarin. Figure 3c contains the normalized response to different sweeteners.

Figure 4 contains intracellular calcium responses in HEK cells stably expressing Gα15, transiently transfected with hT1R2/hT1R3, rT1R2/rT1R3, hT1R2/rT1R3 and rT1R2/hT1R3 in response to 350 mM sucrose, 25 mM tryptophan, 15 mM aspartame, and 0.05 % monellin.

Figure 5 contains the results of a fluorescence plate reactor based assay wherein HEK cells stably expressing Gα15 were transiently transfected with hT1R2 and hT1R3 or hT1R3 alone and contacted with the calcium dye Fluo-4 and a sweet taste stimulus (12.5 mM cyclamate).

Figure 6 contains normalized dose-response curves which show that hT1R2 and hT1R3 function in combination as the human sweet receptor based on their dose-specific interaction with various sweet stimuli (trp, cyclamate, sucrose, neotame, asparame, saccharin and Acek).

Figure 7 contains structural information relating to mGluR1 and T1R1 showing the key ligand binding residues are observed in these molecules.

Figure 8a-8c contains functional data showing HEK cells which stably express Gα15 that are transiently transfected with T1R1/T1R3 respond to glutamate in an intracellular calcium-based assay. Figure 8a shows that intracellular calcium increases in response to increasing glutamate concentration; Figure 8b shows intracellular calcium responds to IMP (2 mM), glutamate (0.5 mM) and 0.2 mM IMP; and Figure 8c shows human T1R1/T1R3 responses for glutamate in the presence and absence of 0.2 mM IMP.

Figures 9a-9b respectively contain the results of an immunofluorescence staining assay using Myc-tagged hT1R2 and a FACS experiment showing that the incorporation of the PDZIP peptide (SEQ ID No: 1) enhanced the expression of a T1R (hT1R2) on the plasma membrane.

Figure 10a through 10b contain calcium imaging data demonstrating that h1TR2/hT1R3 respond to different sweet stimuli.

Figure 11 shows the responses of cell lines which stably express hT1R1/hT1R3 by automated fluorescence imaging to umami taste stimuli.

Figure 12 shows the responses of a cell line which stably expresses hT1R2/hT1R3 by automated fluorescence imaging to sweet taste stimuli.

Figure 13 shows dose-response curves determined using automated fluorescence imaging for a cell line that inducibly expresses the human T1R1/T1R3 taste receptor for L-glutamate in the presence and absence of 0.2mM IMP.

Figures 14 and 15 show the response of a cell line that inducibly expresses the human T1R1/T1R3 taste receptor (I-17 clone) to a panel of L-amino acids. In Figure 14 different C-amino acids at 10mM were tested in the presence and absence of 1 mM IMP. In Figure 15 dose-responses for active amino acids were determined in the presence of 0.2mM IMP.

Figure 16 shows that lactisole inhibits the receptor activities of human T1R2/T1R3 and human T1R1/T1R3.

Figure 17 shows schematics of human-rat T1R chimeras. The chimeras are constructed by fusing the human or rat extracellular domains to the rat or human transmembrane domains respectively, as shown in h2-r2, r2-h2, h3-r3 and r3-h3.

Figure 18 shows neohesperidin dihydrochalcone (NHDC) enhances the activities of T1R1/T1R3 umami taste receptor. [Neohesperidin dihydrochalcone] = 5 μM. The glutamate dose response curve is left-shifted by 2.3 fold (left panel), and the glutamate/IMP dose response is left-shifted by 2.1 fold.

Figure 19 shows that control sweeteners do not affect the activities of T1R1/T1R3 umami taste receptor [Steviocide] = 0.5 mM. [Saccharin] = 1 mM. Glutamate dose response is shown in the left panel, and glutamate/IMP dose response is shown in the right panel.

Figure 20 shows NHDC maps to the transmembrane domain of human T1R3.

Figure 21 shows mapping of a compound to the human T1R2 transmembrane domain.

Figures 22a-d show sweeteners which map to different domains/subunits of the human sweet receptor. Figure 22a shows responses of human and rat sweet receptors to sucrose (200 mM), aspartame (10 mM), neotame (0.1 mM), cyclamate (10 mM), and sucrose (200 mM) in the presence of lactisole (1 mM) (Suc/Lac). HEK-293T cells were transiently transfected with human or rat T1R2, T1R3, and a $G_{\alpha 15}$ chimera $G_{\alpha 15/i1}$, and assayed for intracellular calcium increases in response to sweeteners. Figure 22b shows aspartame and neotame were mapped to N- terminal extracellular domain of human

T1R2. Combinations of T1R chimeras were transiently transfected into HEK-293T cells with $G_{\alpha 15/i1}$, and assayed for responses to sweeteners at the concentrations listed in 23a. The presence or absence of response is what is important. Figure 22c shows cyclamate was mapped to the C-terminal transmembrane domain of human T1R3. Figure 22d shows lactisole was mapped to the transmembrane domain of human T1R3. Different combinations of T1R chimeras were transiently transfected into HEK-293T cells with G $G_{\alpha 15/i1}$, and assayed for responses to sucrose (200 mM) and AceK (10 mM) in the absence or presence of lactisole (1 mM). The activities in B, C and D represent the mean ± SE of number of responding cells for four imaged field of ~1,000 confluent cells.

Figures 23a-d show mutations in T1R2 or T1R3 selectively affect the activity of different sweeteners. Figure 23a shows sequence alignment of the N-terminal ligand binding domain of rat mGluR5 with human and rodent T1R2s. The 8 critical amino acids involved in ligand-binding in mGluR5 are labeled with *, three of the 8 amino acids are conserved in T1R2 and underlined. Figure 23b shows two point mutations in the human T1R2 N-terminal extracellular domain that abolish response to aspartame and neotame without affecting cyclamate. Stable cell lines of hT1R2/hT1R3 (WT), hT1R2 S144A/hT1R3 (S144A) and hT1R2 E302A/hT1R3 (E302A) were generated as describe in the Examples. The dose-responses of these stable lines were determined on FLIPR for sucrose, aspartame, neotame and cyclamate. The activities represent the mean ± SE of fold increases in fluorescence intensities for four recorded wells. Figure 23c shows sequence alignment of human and rodent T1R3 transmembrane domains. The three extracellular loops are underlined and labeled EL1, 2, or 3, according to their order in the protein sequences. Figure 23d shows mutations in the extracellular loop of hT1R3 that abolish response to cyclamate without affecting aspartame. Each of the three extracellular loops of hT1R3 were replaced with rat protein sequence separately, and the resulting hT1R3 mutants were transiently transfected into HEK-293T cells together with $G_{\alpha 15/i1}$, and assayed for responses to sucrose (200mM), aspartame (10 mM) and cyclamate (10 mM). The activities represent the mean ± SE of number of responding cells for four imaged field of ~1,000 confluent cells.

Figures 24a-b show human T1R2 is required for $G_{\alpha 15}$–coupling. Figure 24a shows responses of human, rat and chimeric sweet receptors to sucrose (200 mM) and AceK (10 mM). Stable $G_{\alpha 15}$ cells were transiently transfected with human, rat or

chimeric T1Rs, and assayed for intracellular calcium increases in response to sweeteners. Figure 24b shows $G_{\alpha 15}$–coupling is mediated by human T1R2. The activities represent the mean ± SE of number of responding cells for four imaged field of ~1,000 confluent cells.

Figures 25a-f show the effect of lactisole and cyclamate on the human T1R1/T1R3 umami receptor. Figure 25a shows the response of human T1R1/T1R3 stable cell line to L-glutamate (5 mM) and L- glutamate/IMP (1/0.2 mM) in the absence and presence of lactisole (5 mM). Figure 25b shows the lactisole dose-dependent inhibition curves were determined for L-glutamate (Glu), and L- glutamate with 0.2 mM IMP (Glu/IMP), each at two different concentrations. The IC50s are 0.19 ± 0.02 mM and 0.21 ± 0.01 mM for L-glutamate at 8 and 80 mM; 0.35 ± 0.03 mM and 0.82 ± 0.06 mM for L-glutamate with IMP at 0.8 and 8 mM respectively. Figure 25c shows the dose responses for L-glutamate, with or without 0.2 mM IMP, were determined in the presence of different concentrations of lactisole. In the presence of 0, 25, or 50 µM lactisole, the EC50s are 9.9 ± 1.5 mM, 7.9 ± 0.5 mM, and 7.0 ± 0.3 mM for L-glutamate; in the presence of 0, 100, or 200 µM lactisole, the EC_{50}s are 0.53 ± 0.04 mM, 0.71 ± 0.10 mM, and 0.84 ± 0.10 mM for L-glutamate with IMP. Values represent the mean ± SE for four independent responses. Figure 25d shows the detection thresholds for sweet, umami, and salty taste stimuli were determined in the presence or absence of lactisole. The inhibition effect of lactisole is shown as fold increases in detection thresholds. "Detection thresholds" are defined as the lower limit of detectable tastants. The detection threshold values were averaged over four trials for three subjects. Figure 25e shows the responses of human T1R1/T1R3 stable cell line to threshold level of L-glutamate (4 mM) and endogenous M2 receptor agonist carbachol were assayed on FLIPR in the absence and presence of various concentrations of cyclamate. Figure 25f shows the dose-responses of the human T1R1/T1R3 stable cell line were determined on FLIPR for L-glutamate with or without 0.2 mM IMP in the absence and presence of cyclamate (8 mM). The activities in B, C, E and F represent the mean ± SE of fold increases in fluorescence intensities for four recorded wells. The dose– responses in B, C, E and F were reproduced at least 6 times independently.

Figure 26 shows a working model for the sweet and umami taste receptor structure-function relationships. Filled arrows indicate direct activation, open arrows indicate enhancement, and bar heads indicate inhibition.

Figure 27a shows all 16 combinations of T1Rs and chimeras that were tested for responses to sweeteners and lactisole. rT1R2/T1R3H-R, rT1R2/hT1R3, and T1R2H-R/T1R3R-H show a significant response to cyclamate and they can be inhibited by lactisole. T1R chimeras were transiently transfected into HEK-293T cells with $G_{\alpha 15/i1}$. The activities represent the mean ± SE of number of responding cells for four imaged field of ~1,000 confluent cells, each unit on the Y axis represents 50 responding cells. Abbreviations: Suc (sucrose 100mM); Suc/Lac (sucrose 100mM, lactisole 1 mM); AceK (acesulfame K 10 mM); AceK/Lac (acesulfame K 10 mM, lactisole 1 mM); ATM (aspartame 10mM); NTM (neotame 10 mM); Cyc (cyclamate 10 mM). Figure 27b shows the lactisole dose-dependent inhibition curves of the human sweet receptor were determined for sucrose (Suc), saccharin (Sac), and D-tryptophan (D-Trp), each at two different concentrations. The IC50s are 19.6 ± 0.1 μM and 64.6 ± 0.3 μM for sucrose at 50 mM and 120 mM; 22.6 ± 0.1 μM and 103 ± 7 μM for saccharin at 0.1 and 2 mM; 19.9 ± 0.2 μM and 168 ± 9 μM for D-tryptophan respectively. Figure 27c shows the dose responses of human sweet receptor for sucrose, D-Trp and saccharin were determined with different concentrations of lactisole. In the presence of 0, 10, or 20 μM lactisole, the EC_{50}s are 19.4 ± 0.9 mM, 24.7 ± 1.0 mM, and 31.3 ± 0.3 mM for sucrose; 0.37 ± 0.02 mM, 0.60 ± 0.03 mM, 0.94 ± 0.08 mM for D-Trp; 42 ± 3 μM, 67 ± 6 μM, 118 ± 2 μM for saccharin. Values represent the mean ± SE for four independent responses. The dose-responses in B and C were determined at least 6 times independently, and generated similar results as shown here.

Detailed Description of the Invention

The inventions provides compounds that specifically bind to the wild-type and chimeric sweet and umami taste receptors disclosed herein. Further provided are compounds that specifically bind to the wild-type, chimeric or truncated T1R2 or T1R3 subunits of the sweet and umami receptors.

Binding to the T1R2/T1R3 sweet receptor defines a large genus of molecules. The receptor responds to every sweetener tested, including carbohydrate sugars, amino acids and derivatives, sweet proteins, and synthetic sweeteners. In the meantime, the receptor exhibits stereo-selectivity for certain sweeteners, for example, it responds to D-tryptophan but not L-tryptophan, which is in correlation with taste physiology data.

Thus, the compounds of the invention specifically bind chimeric receptors. Examples include, but are not limited to, a chimeric T1R2/T1R3 receptor comprising a human T1R2 subunit and a rat T1R3 subunit; a chimeric T1R2/T1R3 receptor comprising a rat T1R2 subunit and a human T1R3 subunit; a chimeric T1R2 receptor subunit comprising a human extracellular domain, a rat transmembrane domain and a rat intracellular domain; and a chimeric T1R3 receptor subunit comprising a rat extracellular domain, a human transmembrane domain and a human intracellular domain. The invention provides functional taste receptors, preferably human taste receptors, that are produced by co-expression of a combination of different T1Rs, preferably T1R1/T1R3 or T1R2/T1R3, and the corresponding isolated nucleic acid sequences or fragments, chimeras, or variants thereof that upon co-expression result in a functional taste receptor, i.e., a sweet taste receptor (T1R2/T1R3) or umami taste receptor (T1R1/T1R3).

T1Rs, a family of class C G protein-coupled receptors (GPCRs), are selectively expressed in the taste tissue (Hoon, M.A., et al., Cell, 1999. 96(4): p. 541-51, Bachmanov, A.A., et al., Chem Senses, 2001. 26(7): p. 925-33, Montmayeur, J.P., et al., Nat Neurosci, 2001. 4(5): p. 492-8, Max, M., et al., Nat Genet, 2001. 28(1): p. 58-63, Kitagawa, M., et al., Biochem Biophys Res Commun, 2001. 283(1): p. 236-42 and Nelson, G., et al., Cell, 2001. 106(3): p. 381-90.) Functional expression of T1Rs in HEK293 cells revealed that different combinations of T1Rs respond to sweet and umami taste stimuli (Nelson, G., et al., Cell, 2001. 106(3): p. 381-90, Li, X., et al., Proc Natl Acad Sci U S A, 2002. 99(7): p. 4692-6.) T1R2 and T1R3, when co-expressed in 293 cells, recognize diverse natural and synthetic sweeteners [For the reason mentioned above re "diverse", please consider whether we need this section for enablement. If not, I'd delete. We can discuss], while T1R1 and T1R3 recognize umami taste stimulus L-glutamate, and this response is enhanced by 5'-ribonucleotides, a hallmark of umami taste. Knockout data confirmed that T1Rs indeed mediate mouse sweet and umami tastes (Damak, S., et al., Science, 2003 301(5634): p. 850-3, Zhao, G.Q., et al., Cell 2003 Oct 31;115(3):255-66).

The class C GPCRs possess a large N-terminal extracellular domain, often referred to as the Venus flytrap domain (VFD) (Pin, J.P., Pharmacol Ther, 2003 98(3): p. 325-54), and are known to function as either homodimers, in the cases of metabotropic glutamate receptors (mGluRs) and calcium- sensing receptor (CaR), or

heterodimers, in the case of γ-aminobutyric acid type B receptor ($GABA_BR$). The functional expression data shows a heterodimer mechanism for T1Rs: both T1R1 and T1R2 need to be coexpressed with T1R3 to be functional, which is supported by the overlapping expression patterns of T1Rs in rodent tongue.

It is established herein that T1R family members act in combination with other T1R family members to function as sweet and umami taste receptors. As disclosed in further detail infra in the experimental examples, it has been demonstrated that heterologous cells which co-express hT1R2 and hT1R3 are selectively activated by sweet taste stimuli in a manner that mirrors human sweet taste.

For example, HEK-293-Gα15 cells that co-express hT1R2 and hT1R3 specifically respond to cyclamate, sucrose, aspartame, and saccharin, and the dose responses for these compounds correlate with the psychophysical taste detection thresholds.

Also, as supported by data in the experimental examples, it has been shown that cells which co-express hT1R1 and hT1R3 are selectively activated by glutamate (monosodium glutamate) and 5'-ribonucleotides in a manner that mirrors human umami taste. For example, HEK-293-Gα15 cells that co-express hT1R1 and hT1R3 specifically respond to glutamate and the dose response for this umami-tasting compound correlates with its psychophysical taste detection threshold. Moreover, 5'-ribonucleotides such as IMP enhance the glutamate response of the T1R1/T1R3 receptor, a synergism characteristic of umami taste.

Further, as shown by experimental data in the examples it has been shown that cells which stably and inducibly co-express T1R1/T1R3 selectively respond to the umami taste stimuli L-glutamate and L-aspartate and only weakly respond to other L-amino acids, and at much higher concentrations, providing further evidence that the T1R1/T1R3 receptor can be used in assays to identify compounds that modulate (enhance or block) umami taste stimuli.

Examples of compounds that specifically bind to the sweet receptor and modulate sweet taste can be found in Table 5.

Tables 1-4 provide examples of compounds that specifically bind to the umami receptor and modulate umami taste.

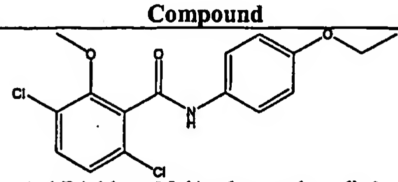
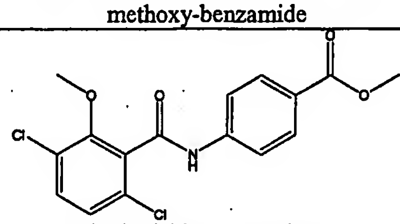
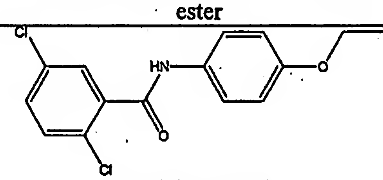
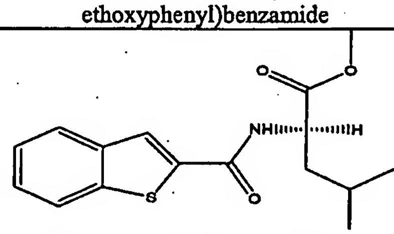
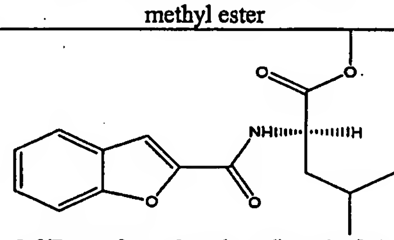
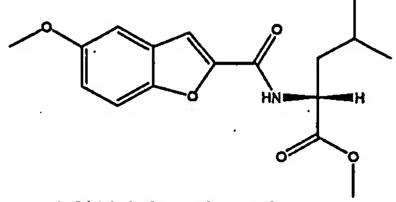
Table 1 - Umami Amides				
Compound No.	Compound	Umami EC_{50} (uM)	Ec_{50} ratio (vs. MSG)	@ (uM)
A1	 3,6-Dichloro-N-(4-ethoxy-phenyl)-2-methoxy-benzamide	0.22	2.74	1
A2	 4-(3,6-Dichloro-2-methoxy-benzoylamino)-benzoic acid methyl ester	0.93	6.98	0.01
A3	 2,5-dichloro-N-(4-ethoxyphenyl)benzamide	1.08	6.14	0.03
A4	 2-[(Benzo[b]thiophene-2-carbonyl)-amino]-4-methyl-pentanoic acid methyl ester	0.4		
A5	 2-[(Benzofuran-2-carbonyl)-amino]-4-methyl-pentanoic acid methyl ester	0.31		
A6	 2-[(5-Methoxy-benzofuran-2-carbonyl)-amino]-4-methyl-pentanoic acid methyl ester	0.32	2.86	1

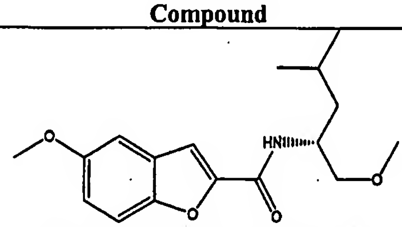
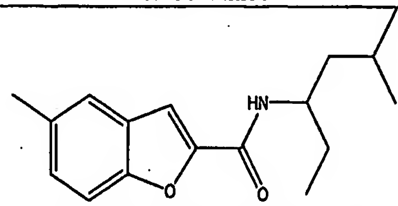
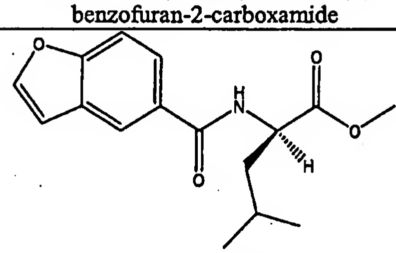
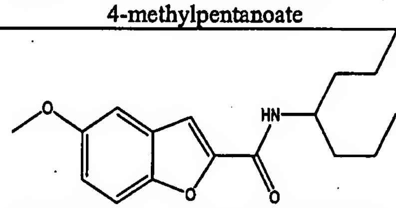
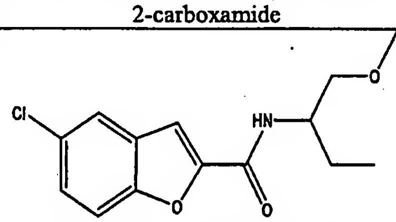
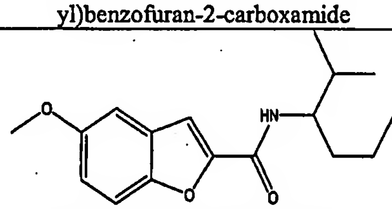
Table 1 - Umami Amides				
Compound No.	Compound	Umami EC_{50} (uM)	Ec_{50} ratio (vs. MSG)	@ (uM)
A7	 (R)-5-methoxy-N-(1-methoxy-4-methylpentan-2-yl)benzofuran-2-carboxamide	0.46		
A8	 5-methyl-N-(5-methylhexan-3-yl) benzofuran-2-carboxamide	0.5		
A9	 2-[(Benzofuran-5-carbonyl)-amino]-4-methyl-pentanoic acid methyl ester(R)-methyl 2-(benzofuran-5-carboxamido)-4-methylpentanoate	0.71		
A10	 N-(heptan-4-yl)-5-methoxybenzofuran-2-carboxamide	0.91	4.51	1
A11	 5-chloro-N-(1-methoxybutan-2-yl)benzofuran-2-carboxamide	1.05	6.5	0.3
A12	 5-methoxy-N-(2-methylhexan-3-yl)benzofuran-2-carboxamide	1.13		

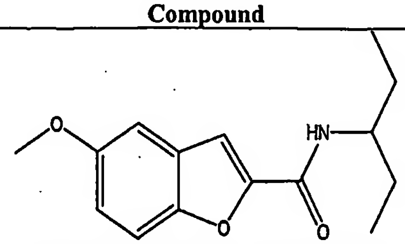
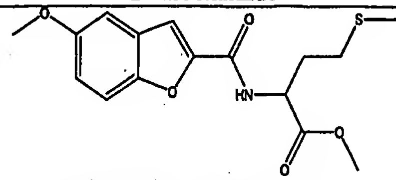
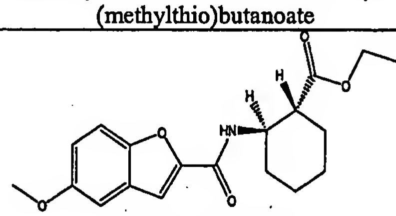
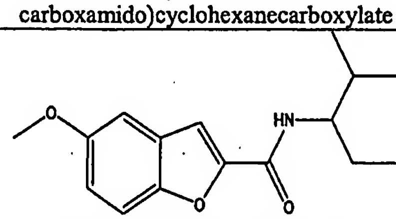
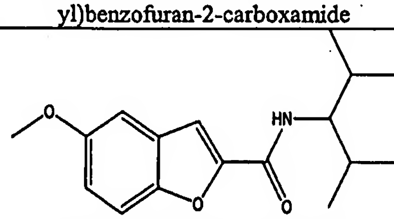
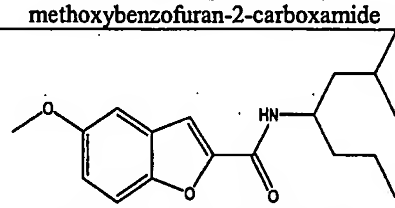
Table 1 - Umami Amides				
Compound No.	**Compound**	**Umami EC_{50} (uM)**	**Ec_{50} ratio (vs. MSG)**	**@ (uM)**
A13	 5-methoxy-N-(pentan-3-yl)benzofuran-2-carboxamide	1.14	4.46	1
A14	 2-[(5-Methoxy-benzofuran-2-carbonyl)-amino]-4-methylsulfanyl-butyric acid methyl ester methyl 2-(5-methoxybenzofuran-2-carboxamido)-4-(methylthio)butanoate	1.14		
A15	 (1R,2R)-ethyl 2-(5-methoxybenzofuran-2-carboxamido)cyclohexanecarboxylate	1.14		
A16	 5-methoxy-N-(2-methylpentan-3-yl)benzofuran-2-carboxamide	1.18		
A17	 N-(2,4-dimethylpentan-3-yl)-5-methoxybenzofuran-2-carboxamide	1.2		
A18	 5-methoxy-N-(2-methylheptan-4-yl)benzofuran-2-carboxamide	1.27		

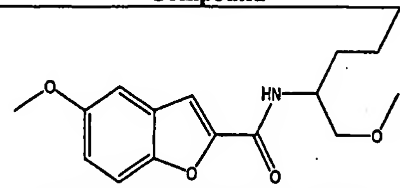
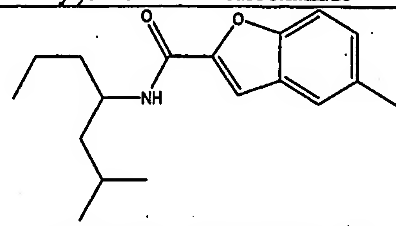
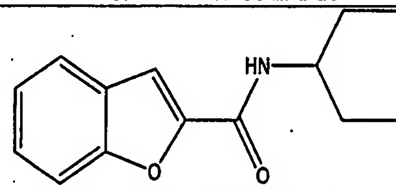
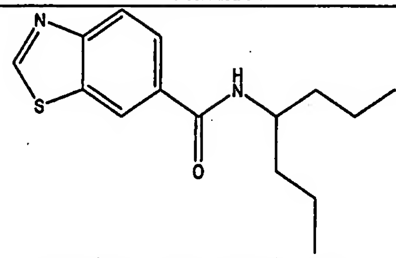
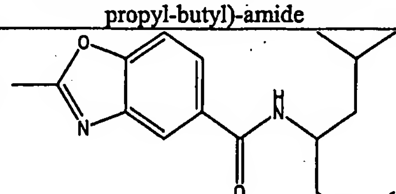
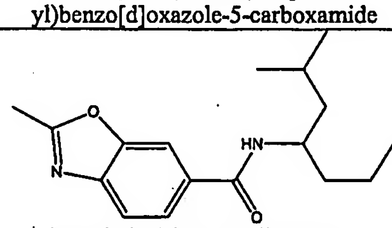
Table 1 - Umami Amides				
Compound No.	Compound	Umami EC$_{50}$ (uM)	Ec$_{50}$ ratio (vs. MSG)	@ (uM)
A19	 5-methoxy-N-(1-methoxypentan-2-yl)benzofuran-2-carboxamide	1.3		
A20	 5-methyl-N-(2-methylheptan-4-yl) benzofuran-2-carboxamide	1.32		
A21	 N-(pentan-3-yl)benzofuran-2-carboxamide	1.52	3.74	1
A22	 Benzothiazole-6-carboxylic acid (1-propyl-butyl)-amide	1.58		
A23	 2-methyl-N-(2-methylheptan-4-yl)benzo[d]oxazole-5-carboxamide	0.38		
A24	 2-methyl-N-(2-methylheptan-4-yl)benzo[d]oxazole-6-carboxamide	1.12		

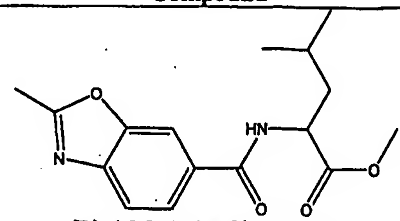
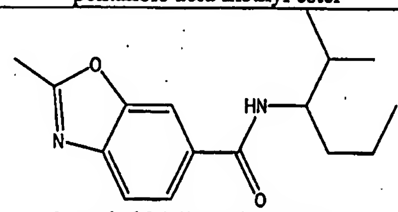
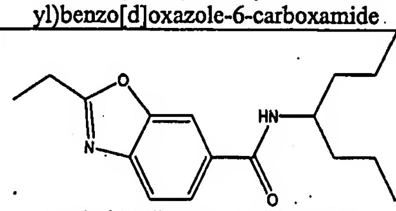
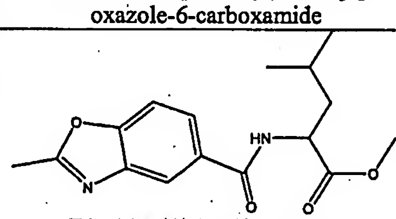
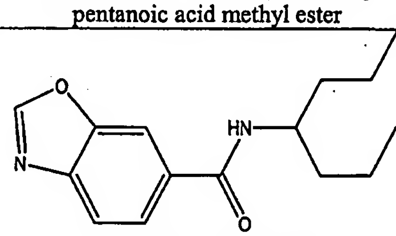
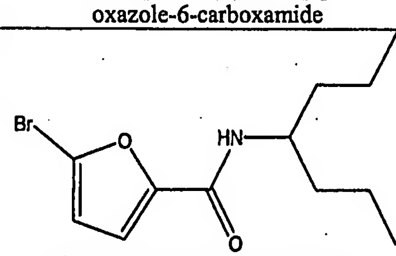
Table 1 - Umami Amides				
Compound No.	Compound	Umami EC_{50} (uM)	Ec_{50} ratio (vs. MSG)	@ (uM)
A25	 (R)-4-Methyl-2-[(2-methyl-benzooxazole-6-carbonyl)-amino]-pentanoic acid methyl ester	1.48		
A26	 2-methyl-N-(2-methylhexan-3-yl)benzo[d]oxazole-6-carboxamide	1.6		
A27	 2-ethyl-N-(heptan-4-yl)benzo[d] oxazole-6-carboxamide	1.61		
A28	 (R)- 4-Methyl-2-[(2-methyl-benzooxazole-5-carbonyl)-amino]-pentanoic acid methyl ester	1.69		
A29	 N-(heptan-4-yl)benzo[d] oxazole-6-carboxamide	1.91		
A30	 5-bromo-N-(heptan-4-yl)furan-2-carboxamide	0.49	12.6	1

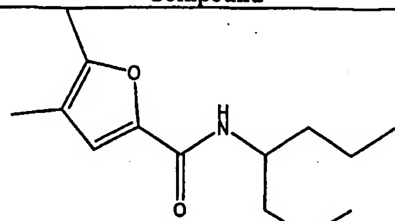
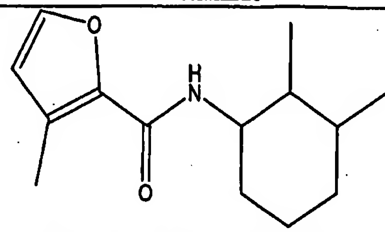
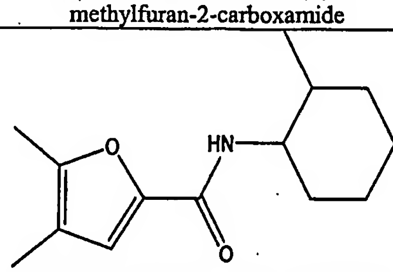
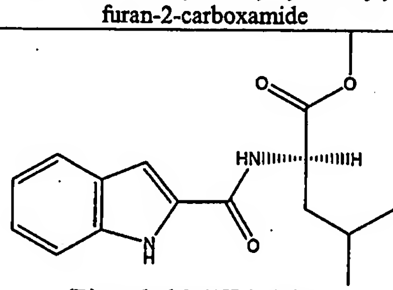
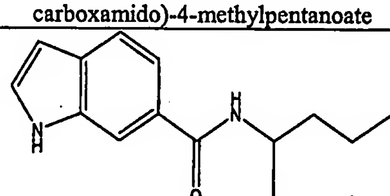
Table 1 - Umami Amides				
Compound No.	Compound	Umami EC_{50} (uM)	Ec_{50} ratio (vs. MSG)	@ (uM)
A31	 N-(heptan-4-yl)-4,5-dimethylfuran-2-carboxamide	0.62	10.04	1
A32	 N-(2,3-dimethylcyclohexyl)-3-methylfuran-2-carboxamide	1.15		
A33	 4,5-dimethyl-N-(2-methylcyclohexyl) furan-2-carboxamide	1.33		
A34	 (R)-methyl 2-(1H-indole-2-carboxamido)-4-methylpentanoate	0.53		
A35	 N-(heptan-4-yl)-1H-indole-6-carboxamide	0.82	8.81	1

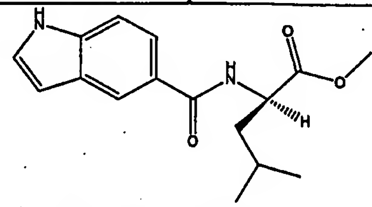
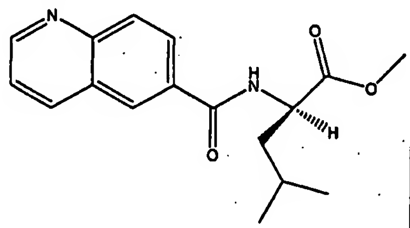
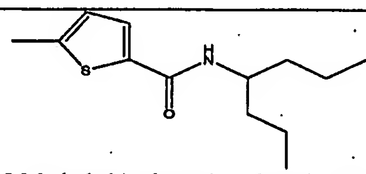
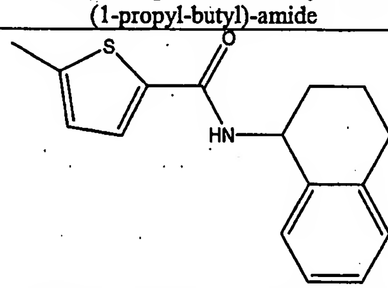
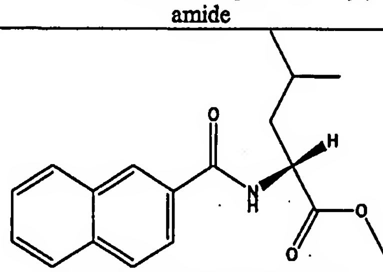
Table 1 - Umami Amides				
Compound No.	Compound	Umami EC_{50} (uM)	Ec_{50} ratio (vs. MSG)	@ (uM)
A36	 (R)-methyl 2-(1H-indole-5-carboxamido)-4-methylpentanoate	1.01		
A37	 (R)-methyl 4-methyl-2-(quinoline-6-carboxamido)pentanoate	1.5		
A38	 5-Methyl-thiophene-2-carboxylic acid (1-propyl-butyl)-amide	1.22	6.54	1
A39	 5-Methyl-thiophene-2-carboxylic acid (1,2,3,4-tetrahydro-naphthalen-1-yl)-amide	1.31	2.3	1
A40	 (R)-methyl 2-(2-naphthamido)-4-methylpentanoate	0.37		

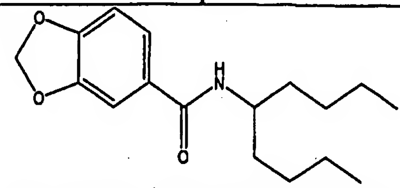
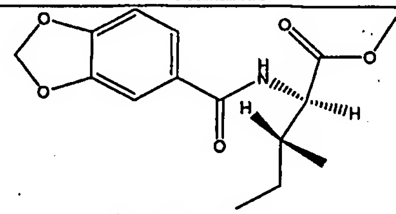
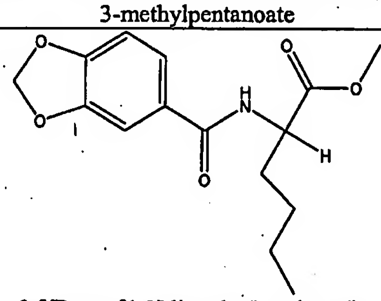
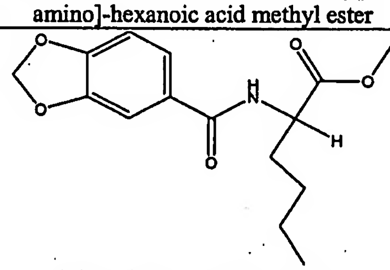
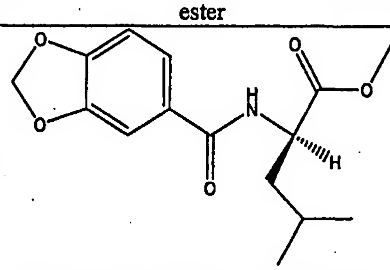
Table 1 - Umami Amides				
Compound No.	Compound	Umami EC50 (uM)	Ec50 ratio (vs. MSG)	@ (uM)
A41	 N-(nonan-5-yl)benzo[d][1,3]dioxole-5-carboxamide	0.7	2.14	3
A42	 (2R,3R)-methyl 2-(benzo[d][1,3]dioxole-5-carboxamido)-3-methylpentanoate	0.35		
A43	 2-[(Benzo[1,3]dioxole-5-carbonyl)-amino]-hexanoic acid methyl ester	0.49		
A44	 (R)-2-[(Benzo[1,3]dioxole-5-carbonyl)-amino]-hexanoic acid methyl ester	0.61		
A45	 (R)-ethyl 2-(benzo[d][1,3]dioxole-5-carboxamido)-4-methylpentanoate	0.88		

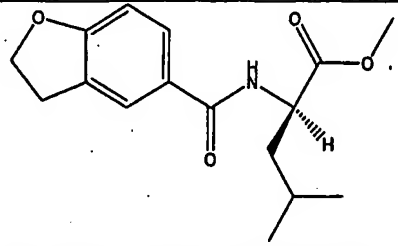
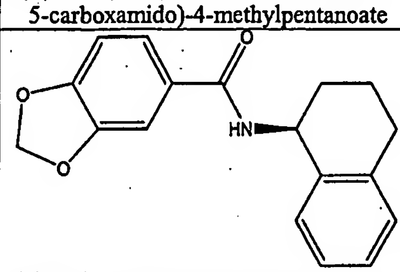
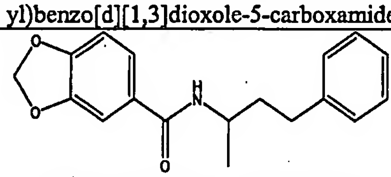
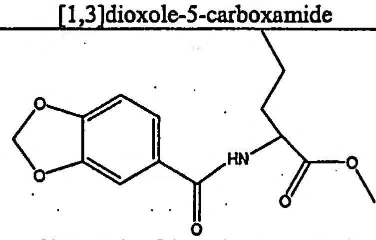
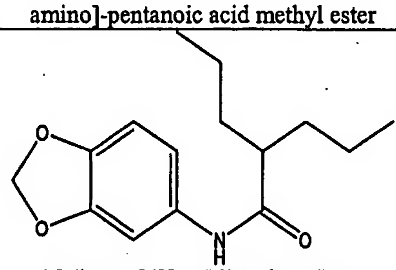
Table 1 - Umami Amides				
Compound No.	Compound	Umami EC_{50} (uM)	Ec_{50} ratio (vs. MSG)	@ (uM)
A46	 (R)-methyl 2-(2,3-dihydrobenzofuran-5-carboxamido)-4-methylpentanoate	1.32		
A47	 (S)-N-(1,2,3,4-tetrahydronaphthalen-1-yl)benzo[d][1,3]dioxole-5-carboxamide	1.33	6.42	0.1
A48	 N-(4-phenylbutan-2-yl)benzo[d][1,3]dioxole-5-carboxamide	1.51	9.27	1
A49	 2-[(Benzo[1,3]dioxole-5-carbonyl)-amino]-pentanoic acid methyl ester	1.54	9.53	1
A50	 N-(benzo[d][1,3]dioxol-5-yl)-2-propylpentanamide	1.57		

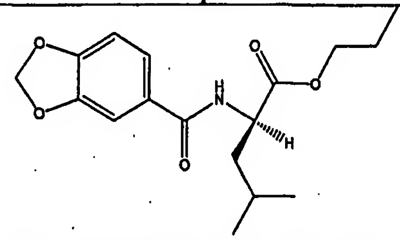
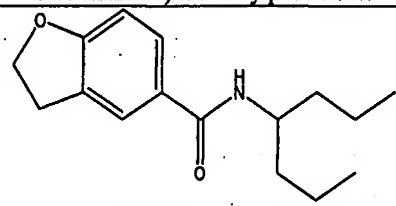
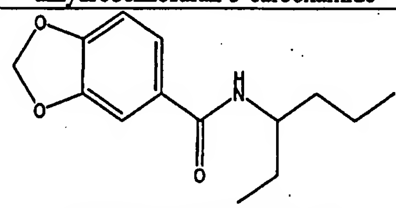
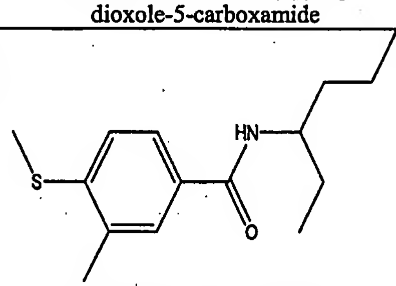
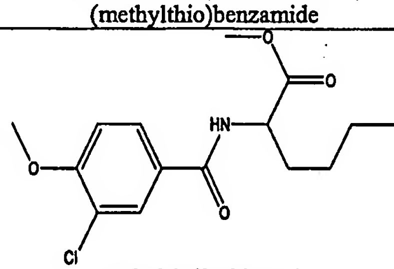
Table 1 - Umami Amides				
Compound No.	Compound	Umami EC_{50} (uM)	Ec_{50} ratio (vs. MSG)	@ (uM)
A51	 (R)-propyl 2-(benzo[d][1,3]dioxole-5-carboxamido)-4-methylpentanoate	1.58		
A52	 N-(heptan-4-yl)-2,3-dihydrobenzofuran-5-carboxamide	1.65		
A53	 N-(hexan-3-yl)benzo[d][1,3] dioxole-5-carboxamide	1.83		
A54	 N-(hexan-3-yl)-3-methyl-4-(methylthio)benzamide	0.12		
A55	 methyl 2-(3-chloro-4-methoxybenzamido)hexanoate	0.12		

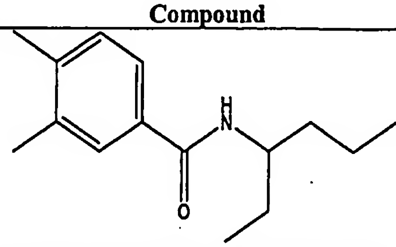
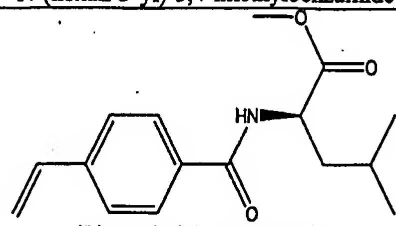
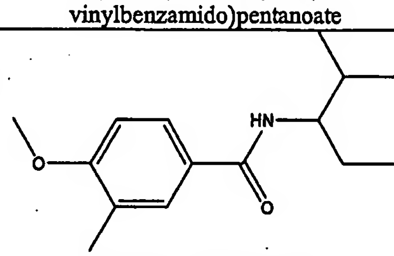
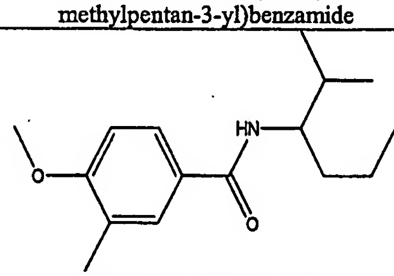
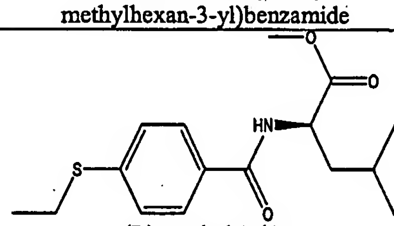
Table 1 - Umami Amides				
Compound No.	Compound	Umami EC_{50} (uM)	Ec_{50} ratio (vs. MSG)	@ (uM)
A56	 N-(hexan-3-yl)-3,4-imethylbenzamide	0.14		
A57	 (R)-methyl 4-methyl-2-(4-vinylbenzamido)pentanoate	0.18		
A58	 4-methoxy-3-methyl-N-(2-methylpentan-3-yl)benzamide	0.2		
A59	 4-methoxy-3-methyl-N-(2-methylhexan-3-yl)benzamide	0.2		
A60	 (R)-methyl 2-(4-(ethylthio)benzamido)-4-methylpentanoate	0.2		

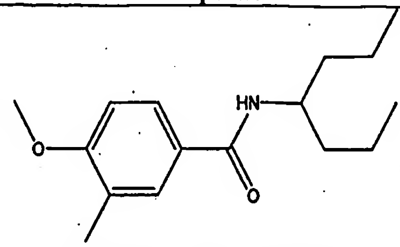
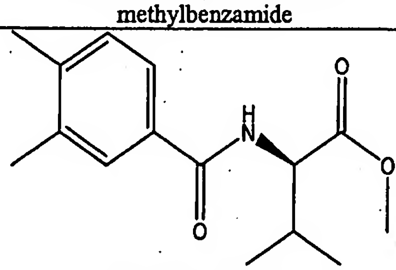
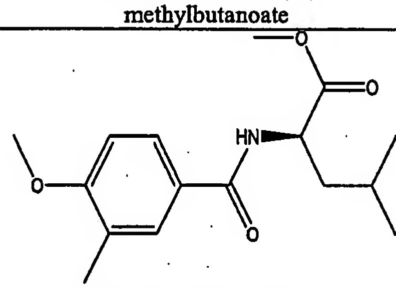
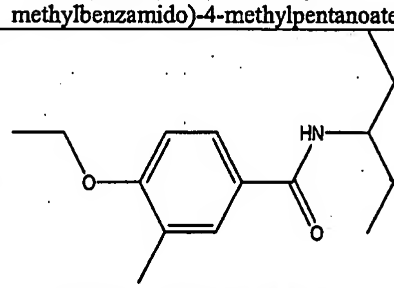
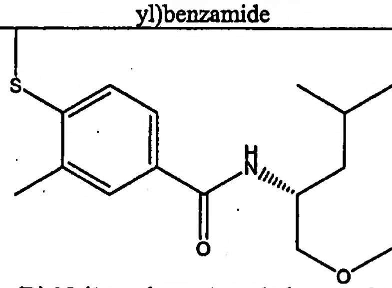
Table 1 - Umami Amides				
Compound No.	Compound	Umami EC_{50} (uM)	Ec_{50} ratio (vs. MSG)	@ (uM)
A61	 N-(heptan-4-yl)-4-methoxy-3-methylbenzamide	0.22		
A62	 (R)-methyl 2-(3,4-dimethylbenzamido)-3-methylbutanoate	0.25		
A63	 (R)-methyl 2-(4-methoxy-3-methylbenzamido)-4-methylpentanoate	0.25		
A64	 4-ethoxy-3-methyl-N-(pentan-3-yl)benzamide	0.26		
A65	 (R)-N-(1-methoxy-4-methylpentan-2-yl)-3-methyl-4-(methylthio)benzamide	0.29		

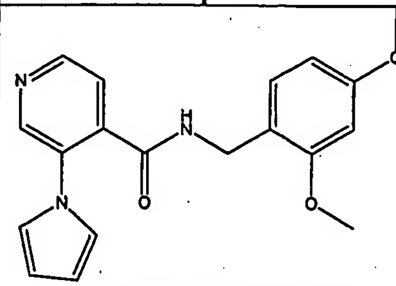
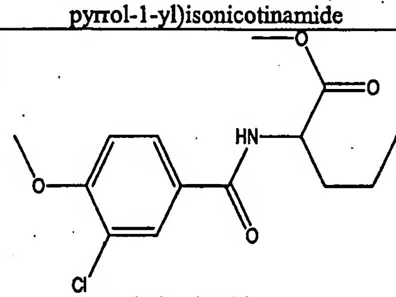
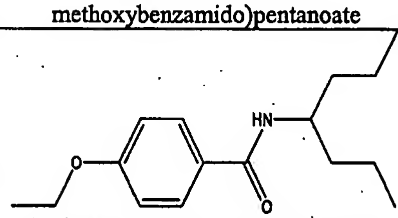
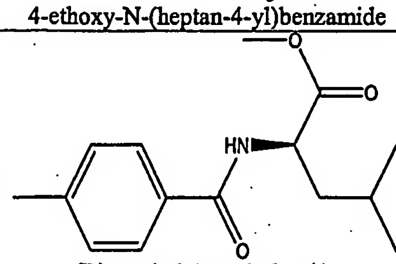
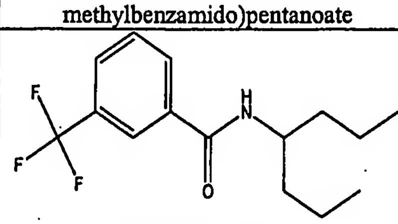
Table 1 - Umami Amides				
Compound No.	Compound	Umami EC$_{50}$ (uM)	Ec$_{50}$ ratio (vs. MSG)	@ (uM)
A66	 N-(2,4-dimethoxybenzyl)-3-(1H-pyrrol-1-yl)isonicotinamide	0.29		
A67	 methyl 2-(3-chloro-4-methoxybenzamido)pentanoate	0.29	10.75	1
A68	 4-ethoxy-N-(heptan-4-yl)benzamide	0.32	2.62	0.3
A69	 (R)-methyl 4-methyl-2-(4-methylbenzamido)pentanoate	0.32		
A70	 N-(heptan-4-yl)-3-(trifluoromethyl)benzamide	0.33		

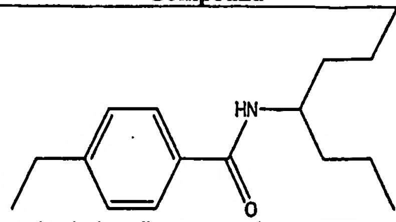
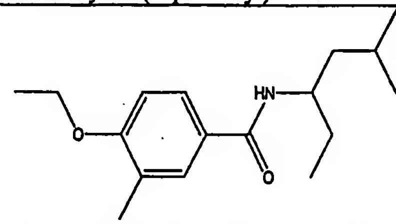
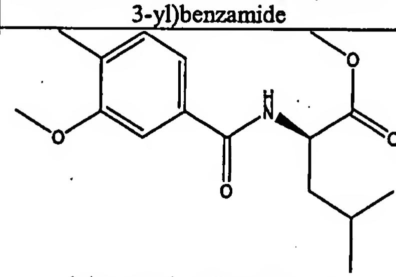
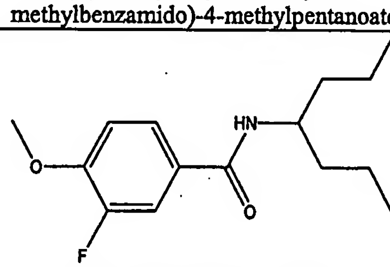
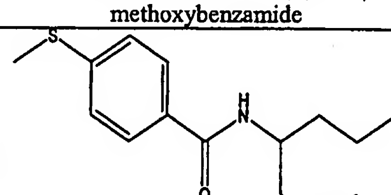
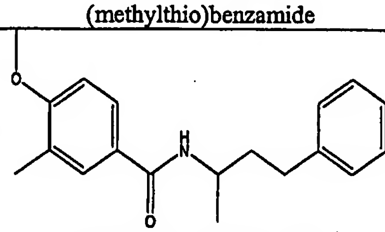
Table 1 - Umami Amides				
Compound No.	Compound	Umami EC_{50} (uM)	Ec_{50} ratio (vs. MSG)	@ (uM)
A71	 4-ethyl-N-(heptan-4-yl)benzamide	0.34		
A72	 4-ethoxy-3-methyl-N-(5-methylhexan-3-yl)benzamide	0.34		
A73	 (R)-methyl 2-(3-methoxy-4-methylbenzamido)-4-methylpentanoate	0.34		
A74	 3-fluoro-N-(heptan-4-yl)-4-methoxybenzamide	0.35	4.98	0.3
A75	 N-(heptan-4-yl)-4-(methylthio)benzamide	0.39		
A76	 4-methoxy-3-methyl-N-(4-phenylbutan-2-yl)benzamide	0.4		

Table 1 - Umami Amides

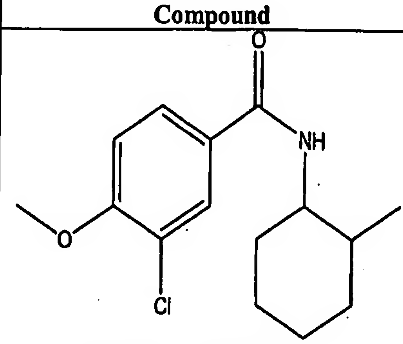
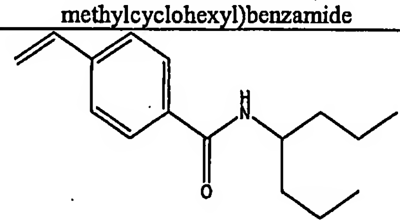
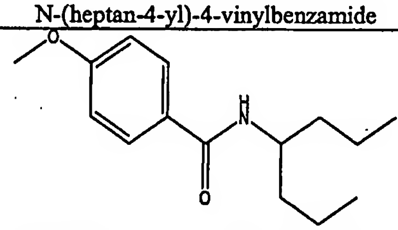
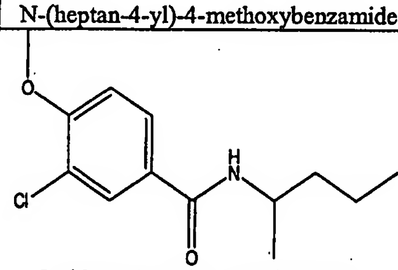
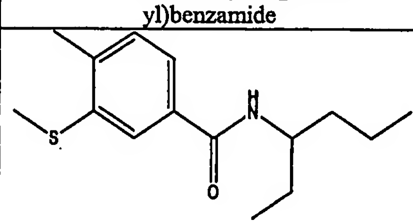
Compound No.	Compound	Umami EC_{50} (uM)	Ec_{50} ratio (vs. MSG)	@ (uM)
A77	 3-chloro-4-methoxy-N-(2-methylcyclohexyl)benzamide	0.44		
A78	 c N-(heptan-4-yl)-4-vinylbenzamide	0.46	10.22	0.3
A79	 N-(heptan-4-yl)-4-methoxybenzamide	0.46		
A80	 3-chloro-4-methoxy-N-(pentan-2-yl)benzamide	0.47	5.12	0.1
A81	 N-(hexan-3-yl)-4-methyl-3-(methylthio)benzamide	0.5		

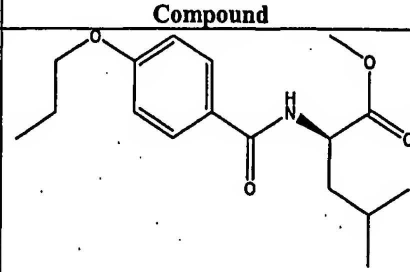
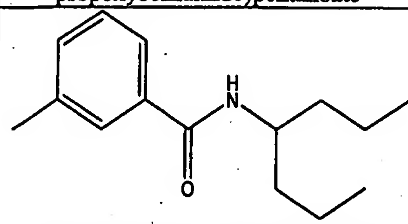
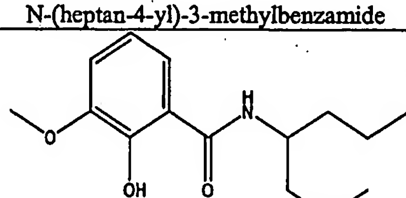
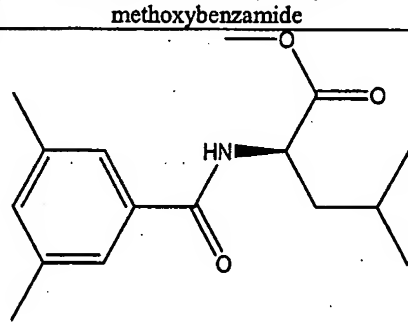
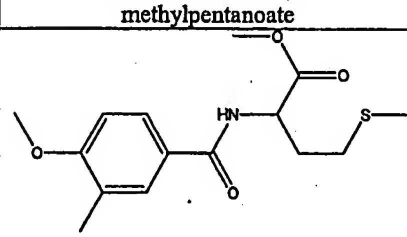
Table 1 - Umami Amides				
Compound No.	Compound	Umami EC_{50} (uM)	Ec_{50} ratio (vs. MSG)	@ (uM)
A82	 (R)-methyl 4-methyl-2-(4-propoxybenzamido)pentanoate	0.51		
A83	 N-(heptan-4-yl)-3-methylbenzamide	0.52		
A84	 N-(heptan-4-yl)-2-hydroxy-3-methoxybenzamide	0.53		
A85	 (R)-methyl 2-(3,5-dimethylbenzamido)-4-methylpentanoate	0.53		
A86	 methyl 2-(4-methoxy-3-methylbenzamido)-4-(methylthio)butanoate	0.53		

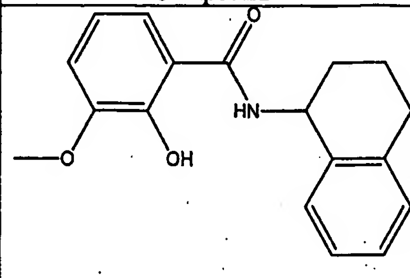
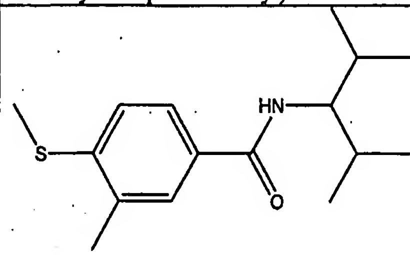
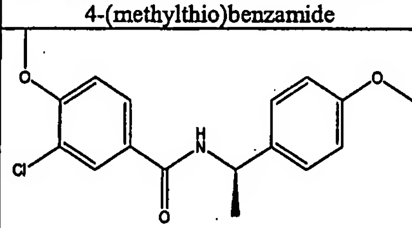
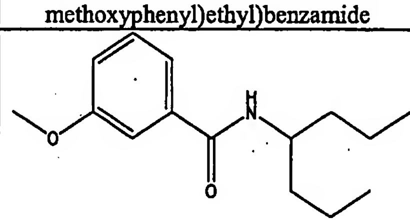
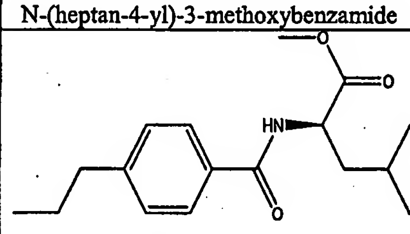
Table 1 - Umami Amides				
Compound No.	Compound	Umami EC_{50} (uM)	Ec_{50} ratio (vs. MSG)	@ (uM)
A87	 2-hydroxy-3-methoxy-N-(1,2,3,4-tetrahydronaphthalen-1-yl)benzamide	0.54	3.8	1
A88	 N-(2,4-dimethylpentan-3-yl)-3-methyl-4-(methylthio)benzamide	0.55		
A89	 (R)-3-chloro-4-methoxy-N-(1-(4-methoxyphenyl)ethyl)benzamide	0.6	2.85	1
A90	 N-(heptan-4-yl)-3-methoxybenzamide	0.61		
A91	 (R)-methyl 4-methyl-2-(4-propylbenzamido)pentanoate	0.62		

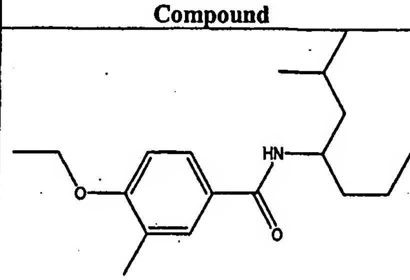
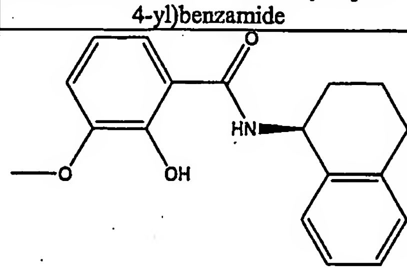
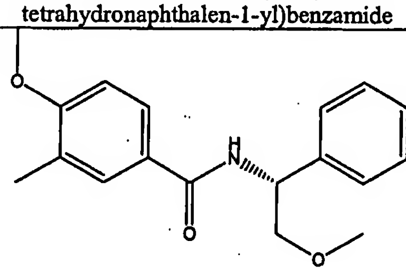
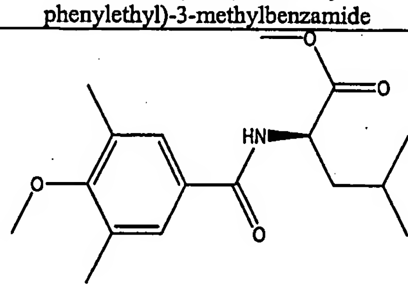
Table 1 - Umami Amides				
Compound No.	Compound	Umami EC_{50} (uM)	Ec_{50} ratio (vs. MSG)	@ (uM)
A92	 4-ethoxy-3-methyl-N-(2-methylheptan-4-yl)benzamide	0.65		
A93	 (S)-2-hydroxy-3-methoxy-N-(1,2,3,4-tetrahydronaphthalen-1-yl)benzamide	0.7	5.7	1
A94	 (R)-4-methoxy-N-(2-methoxy-1-phenylethyl)-3-methylbenzamide	0.72		
A95	 (R)-methyl 2-(4-methoxy-3,5-dimethylbenzamido)-4-methylpentanoate	0.74		

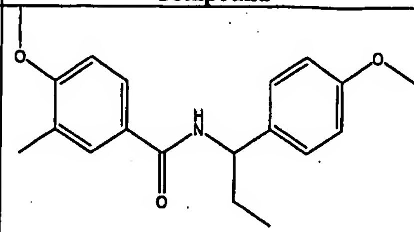
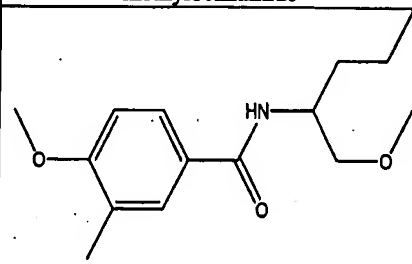
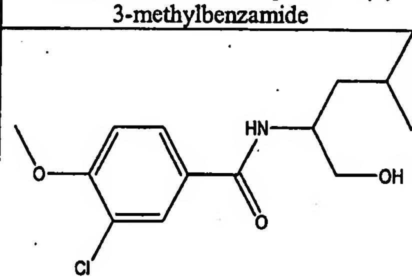
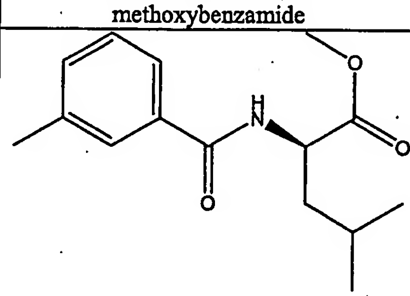
Table 1 - Umami Amides				
Compound No.	Compound	Umami EC_{50} (uM)	Ec_{50} ratio (vs. MSG)	@ (uM)
A96	 4-methoxy-N-(1-(4-methoxyphenyl)propyl)-3-methylbenzamide	0.76		
A97	 4-methoxy-N-(1-methoxypentan-2-yl)-3-methylbenzamide	0.85		
A98	 3-chloro-N-(1-hydroxy-4-methylpentan-2-yl)-4-methoxybenzamide	0.88		
A99	 (R)-methyl 4-methyl-2-(3-methylbenzamido)pentanoate	0.89		

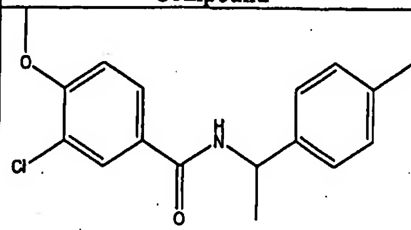
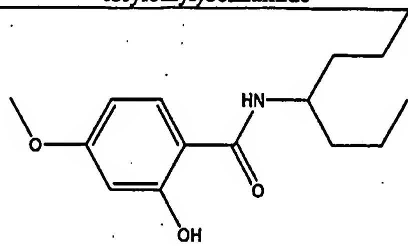
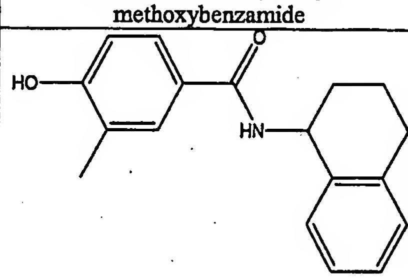
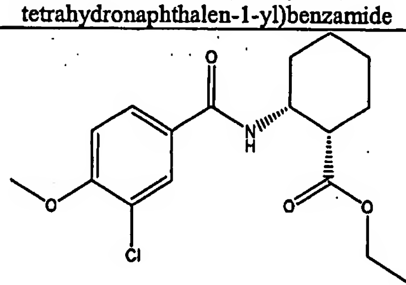
Table 1 - Umami Amides				
Compound No.	Compound	Umami EC_{50} (uM)	Ec_{50} ratio (vs. MSG)	@ (uM)
A100	 3-chloro-4-methoxy-N-(1-p-tolylethyl)benzamide	1.1		
A101	 N-(heptan-4-yl)-2-hydroxy-4-methoxybenzamide	1.16	7.62	1
A102	 4-hydroxy-3-methyl-N-(1,2,3,4-tetrahydronaphthalen-1-yl)benzamide	1.32	9.49	1
A103	 (1S,2R)-ethyl 2-(3-chloro-4-methoxybenzamido) cyclohexanecarboxylate	1.36		

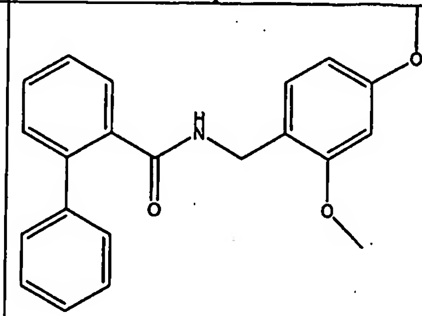
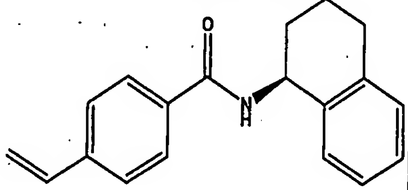
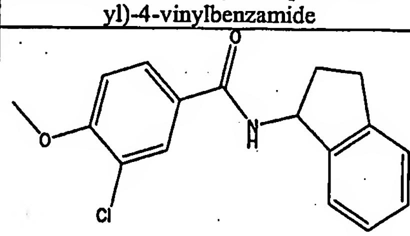
Table 1 - Umami Amides				
Compound No.	Compound	Umami EC_{50} (uM)	Ec_{50} ratio (vs. MSG)	@ (uM)
A104	 Biphenyl-2-carboxylic acid 2,4-dimethoxy-benzylamide	1.37		
A105	 (S)-N-(1,2,3,4-tetrahydronaphthalen-1-yl)-4-vinylbenzamide	1.38	2.79	1
A106	 3-chloro-N-(2,3-dihydro-1H-inden-1-yl)-4-methoxybenzamide	1.39	4.01	0.3

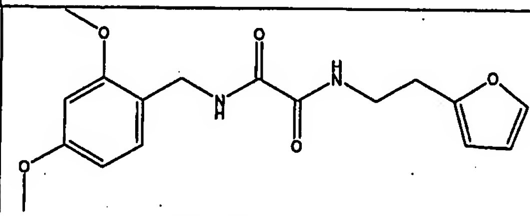
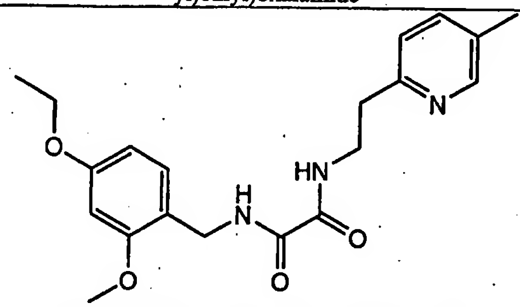
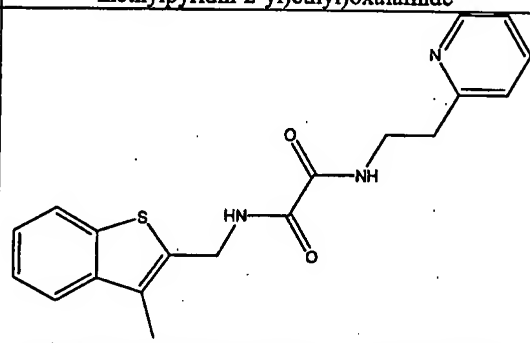
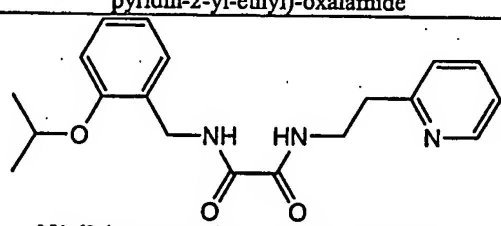
Table 2 - Umami Oxalamides			
Compound No.	Compound	Umami EC_{50} (uM)	Ec_{50} ratio (vs. MSG)
B1	 N1-(2,4-dimethoxybenzyl)-N2-(2-(furan-2-yl)ethyl)oxalamide	0.18	
B2	 N1-(4-ethoxy-2-methoxybenzyl)-N2-(2-(5-methylpyridin-2-yl)ethyl)oxalamide	0.19	
B3	 N-(3-Methyl-benzo[b]thiophen-2-ylmethyl)-N'-(2-pyridin-2-yl-ethyl)-oxalamide	0.81	
B4	 N1-(2-isopropoxybenzyl)-N2-(2-(pyridin-2-yl)ethyl)oxalamide	1.22	

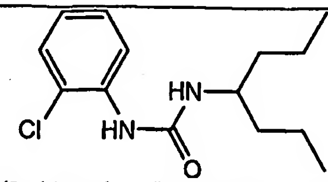
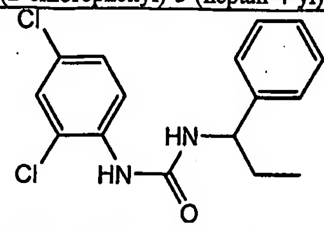
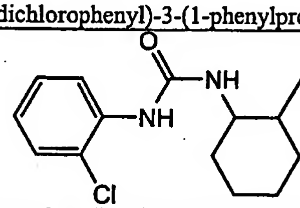
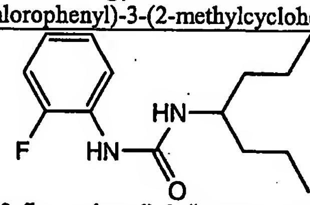
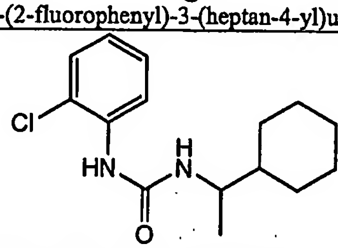
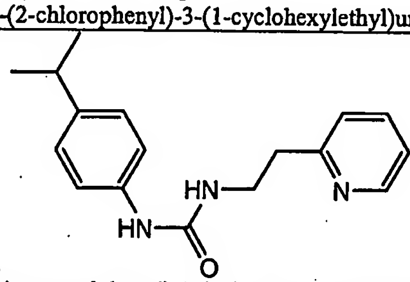
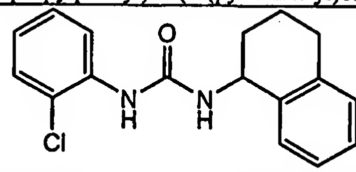
Table 3 - Umami Ureas				
Compound No.	IUPAC Name	Umami EC_{50} uM	Ec50 ratio (vs. MSG)	Con. (uM)
C1	 1-(2-chlorophenyl)-3-(heptan-4-yl)urea	0.37	4.95	1
C2	 1-(2,4-dichlorophenyl)-3-(1-phenylpropyl)urea	0.49	4.52	1
C3	 1-(2-chlorophenyl)-3-(2-methylcyclohexyl)urea	0.52	3.24	3
C4	 1-(2-fluorophenyl)-3-(heptan-4-yl)urea	0.79	12.15	3
C5	 1-(2-chlorophenyl)-3-(1-cyclohexylethyl)urea	0.84	9.08	1
C6	 1-(4-isopropylphenyl)-3-(2-(pyridin-2-yl)ethyl)urea	0.98		
C7	 1-(2-chlorophenyl)-3-(1,2,3,4-tetrahydronaphthalen-1-yl)urea	0.99	3.68	1

Table 3 - Umami Ureas

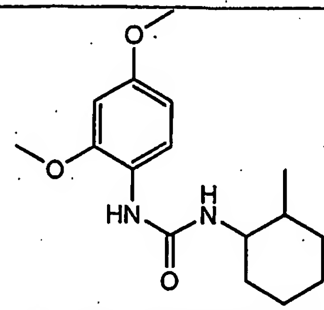
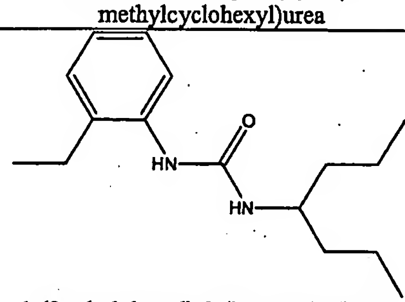
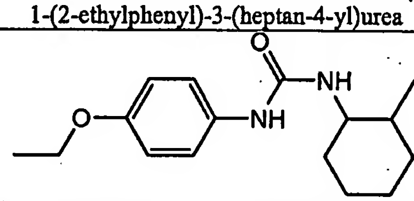
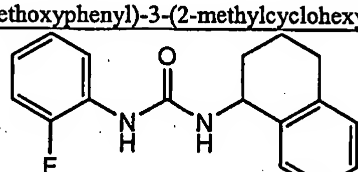
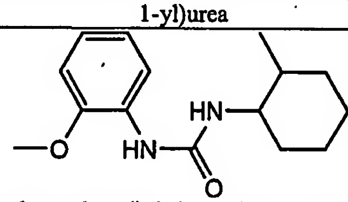
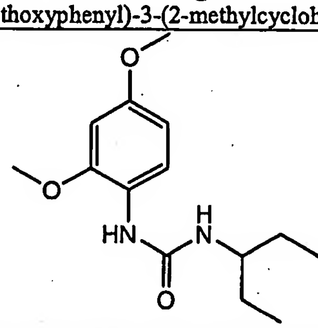
Compound No.	IUPAC Name	Umami EC_{50} uM	Ec50 ratio (vs. MSG)	Con. (uM)
C8	 1-(2,4-dimethoxyphenyl)-3-(2-methylcyclohexyl)urea	1.41	2.62	0.3
C9	 1-(2-ethylphenyl)-3-(heptan-4-yl)urea	1.42		
C10	 1-(4-ethoxyphenyl)-3-(2-methylcyclohexyl)urea	1.51	2.1	0.3
C11	 1-(2-fluorophenyl)-3-(1,2,3,4-tetrahydronaphthalen-1-yl)urea	1.65	4.49	1
C12	 1-(2-methoxyphenyl)-3-(2-methylcyclohexyl)urea	1.67		
C13	 1-(2,4-dimethoxyphenyl)-3-(pentan-3-yl)urea	1.72	11.87	1

Table 4 - Umami Acrylamides

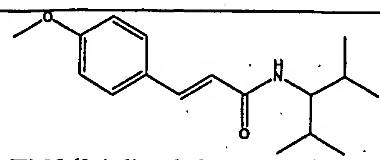
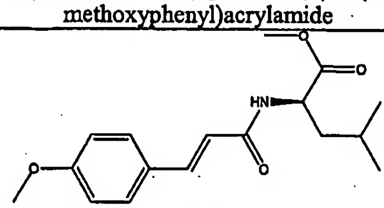
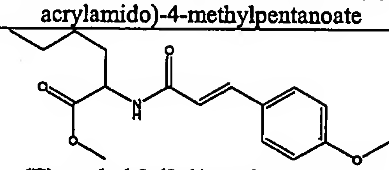
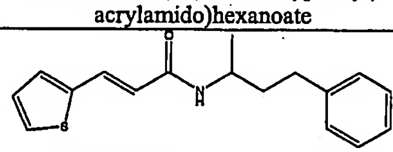
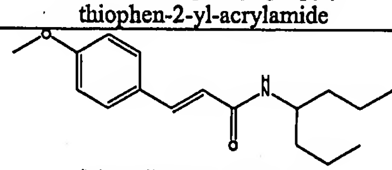
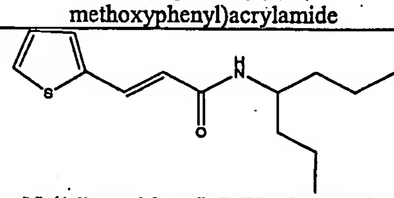
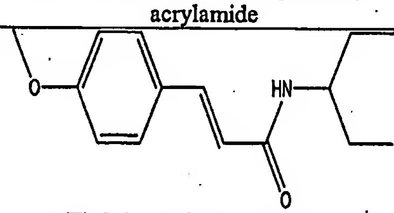
Compound No.	Compound	Umami EC_{50} (uM)	Ec_{50} ratio (vs. MSG)	@ (uM)
D1	 (E)-N-(2,4-dimethylpentan-3-yl)-3-(4-methoxyphenyl)acrylamide	0.29	3.46	1
D2	 (R,E)-methyl 2-(3-(4-methoxyphenyl) acrylamido)-4-methylpentanoate	0.32		
D3	 (E)-methyl 2-(3-(4-methoxyphenyl) acrylamido)hexanoate	0.63		
D4	 N-(1-Methyl-3-phenyl-propyl)-3-thiophen-2-yl-acrylamide	0.69	9.73	1
D5	 (E)-N-(heptan-4-yl)-3-(4-methoxyphenyl)acrylamide	0.72	3.48	0.3
D6	 N-(1-Propyl-butyl)-3-thiophen-2-yl-acrylamide	0.75	6.3	1
D7	 (E)-3-(4-methoxyphenyl)-N-(pentan-3-yl)acrylamide	0.82	9.62	1

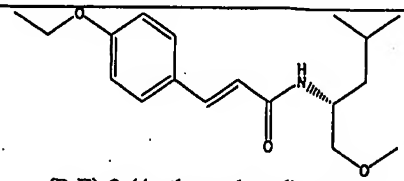
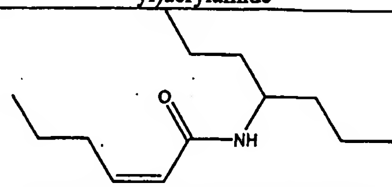
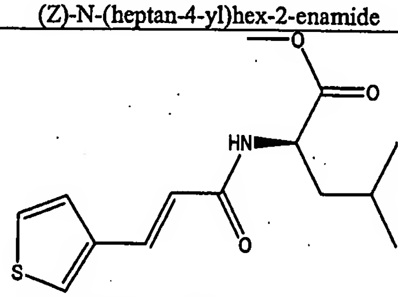
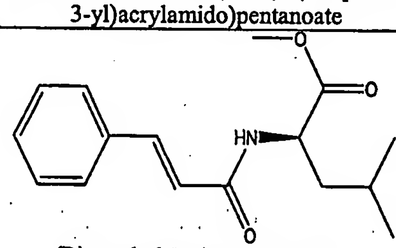
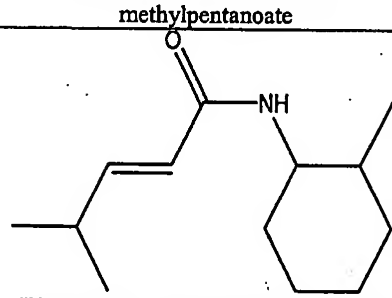
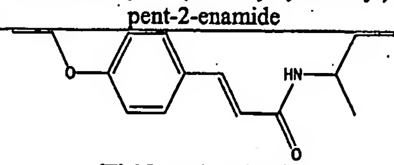
Table 4 - Umami Acrylamides				
Compound No.	Compound	Umami EC_{50} (uM)	Ec_{50} ratio (vs. MSG)	@ (uM)
D8	 (R,E)-3-(4-ethoxyphenyl)-N-(1-methoxy-4-methylpentan-2-yl)acrylamide	0.94		
D9	 (Z)-N-(heptan-4-yl)hex-2-enamide	0.98		
D10	 (R,E)-methyl 4-methyl-2-(3-(thiophen-3-yl)acrylamido)pentanoate	1.09		
D11	 (R)-methyl 2-cinnamamido-4-methylpentanoate	1.17		
D12	 (E)-4-methyl-N-(2-methylcyclohexyl) pent-2-enamide	1.28		
D13	 (E)-N-sec-butyl-3-(4-ethoxyphenyl)acrylamide	1.31	2.7	0.3

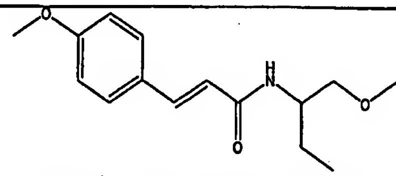
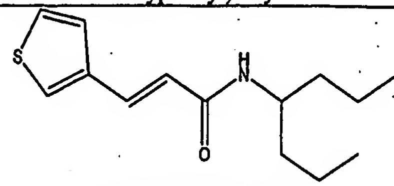
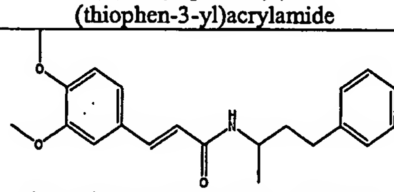
Table 4 - Umami Acrylamides				
Compound No.	**Compound**	**Umami EC_{50} (uM)**	**Ec_{50} ratio (vs. MSG)**	**@ (uM)**
D14	 (E)-N-(1-methoxybutan-2-yl)-3-(4-methoxyphenyl)acrylamide	1.43	8.48	1
D15	 (E)-N-(heptan-4-yl)-3-(thiophen-3-yl)acrylamide	1.54	2.22	0.3
D16	 (E)-3-(3,4-dimethoxyphenyl)-N-(4-phenylbutan-2-yl)acrylamide	1.56	3.13	1

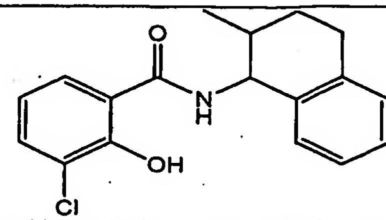
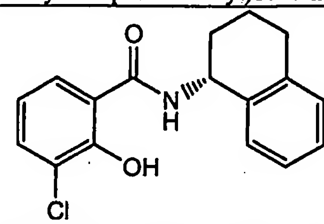
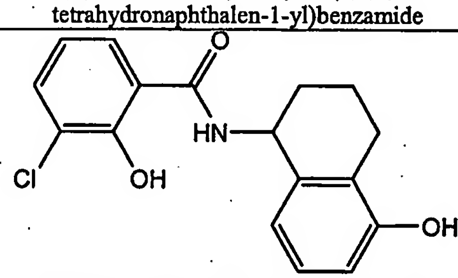
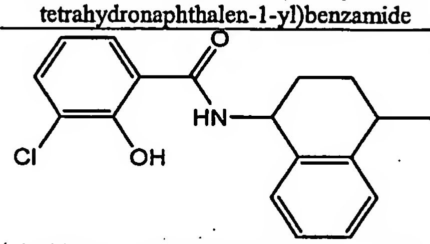
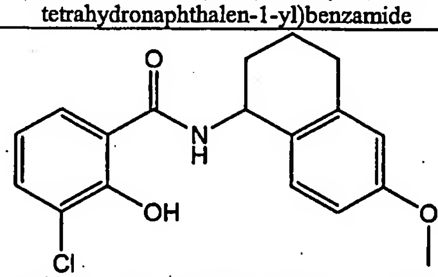
Table 5 - Sweet EnhancerAmides				
Compound No.	Compound	Sweet EC_{50} uM	Umami EC_{50} uM	Umami EC_{50} ratio
E1	 3-chloro-2-hydroxy-N-(2-methyl-1,2,3,4-tetrahydronaphthalen-1-yl)benzamide	0.19		
E2	 (R)-3-chloro-2-hydroxy-N-(1,2,3,4-tetrahydronaphthalen-1-yl)benzamide	0.65		
E3	 3-chloro-2-hydroxy-N-(5-hydroxy-1,2,3,4-tetrahydronaphthalen-1-yl)benzamide	1.03		
E4	 3-chloro-2-hydroxy-N-(4-methyl-1,2,3,4-tetrahydronaphthalen-1-yl)benzamide	1.61		
E5	 3-chloro-2-hydroxy-N-(6-methoxy-1,2,3,4-tetrahydronaphthalen-1-yl)benzamide	1.61		

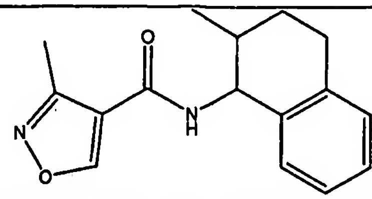
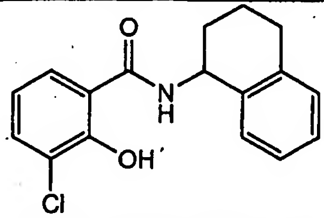
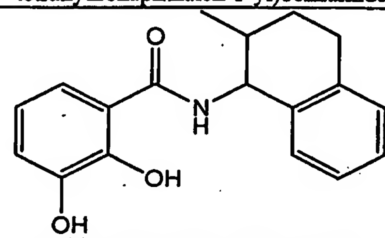
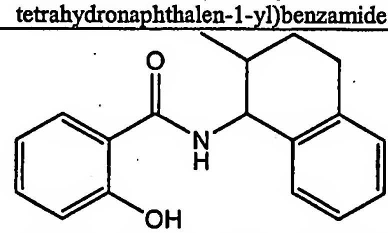
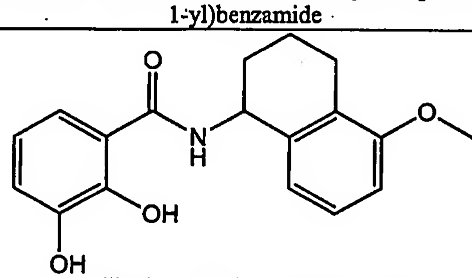
Table 5 - Sweet EnhancerAmides				
Compound No.	**Compound**	**Sweet EC_{50} uM**	**Umami EC_{50} uM**	**Umami EC_{50} ratio**
E6	 3-methyl-N-(2-methyl-1,2,3,4-tetrahydronaphthalen-1-yl)isoxazole-4-carboxamide	1.48		
E7	 3-chloro-2-hydroxy-N-(1,2,3,4-tetrahydronaphthalen-1-yl)benzamide	1.81		4.04
E8	 2,3-dihydroxy-N-(2-methyl-1,2,3,4-tetrahydronaphthalen-1-yl)benzamide	1.98		
E9	 2-hydroxy-N-(2-methyl-1,2,3,4-tetrahydronaphthalen-1-yl)benzamide	2.36		
E10	 2,3-dihydroxy-N-(5-methoxy-1,2,3,4-tetrahydronaphthalen-1-yl)benzamide	2.44		

Table 5 - Sweet EnhancerAmides

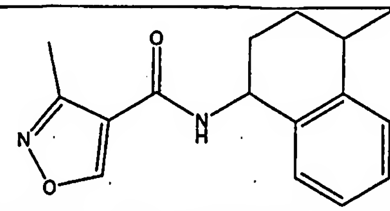
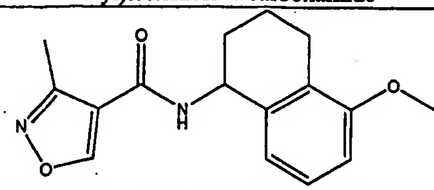
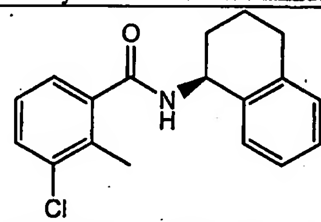
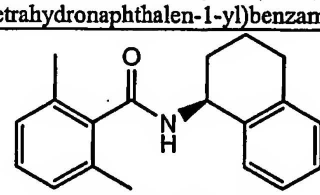
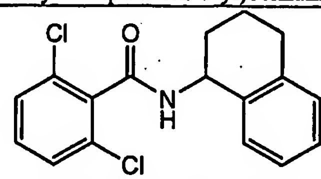
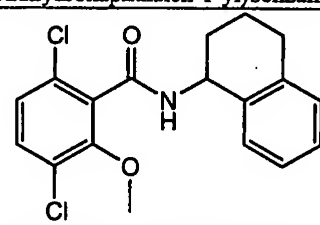
Compound No.	Compound	Sweet EC_{50} uM	Umami EC_{50} uM	Umami EC_{50} ratio
E11	 3-methyl-N-(4-methyl-1,2,3,4-tetrahydronaphthalen-1-yl)isoxazole-4-carboxamide	2.46		
E12	 N-(5-methoxy-1,2,3,4-tetrahydronaphthalen-1-yl)-3-methylisoxazole-4-carboxamide	2.85		
E13	 (S)-3-chloro-2-methyl-N-(1,2,3,4-tetrahydronaphthalen-1-yl)benzamide	2.91		
E14	 (S)-2,6-dimethyl-N-(1,2,3,4-tetrahydronaphthalen-1-yl)benzamide	2.91		
E15	 2,6-dichloro-N-(1,2,3,4-tetrahydronaphthalen-1-yl)benzamide	3.02		
E16	 3,6-dichloro-2-methoxy-N-(1,2,3,4-tetrahydronaphthalen-1-yl)benzamide	3.04		

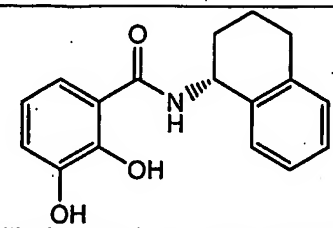
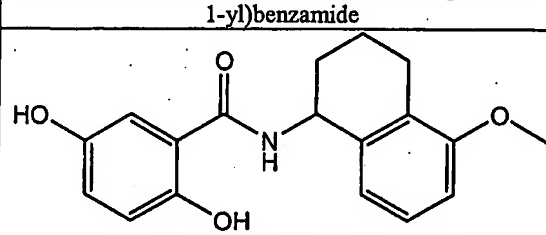
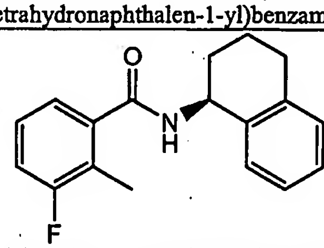
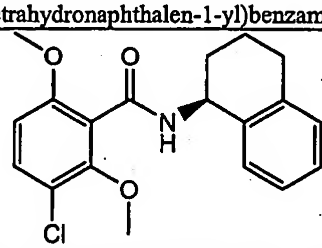
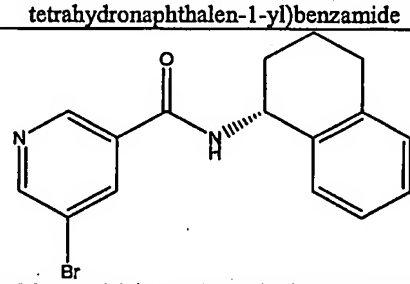
Table 5 - Sweet EnhancerAmides				
Compound No.	**Compound**	**Sweet EC_{50} uM**	**Umami EC_{50} uM**	**Umami EC_{50} ratio**
E17	 (R)-2,3-dihydroxy-N-(1,2,3,4-tetrahydronaphthalen-1-yl)benzamide	3.13		
E18	 2,5-dihydroxy-N-(5-methoxy-1,2,3,4-tetrahydronaphthalen-1-yl)benzamide	3.38		
E19	 (S)-3-fluoro-2-methyl-N-(1,2,3,4-tetrahydronaphthalen-1-yl)benzamide	3.57		
E20	 (S)-3-chloro-2,6-dimethoxy-N-(1,2,3,4-tetrahydronaphthalen-1-yl)benzamide	4.13		
E21	 (R)-5-bromo-N-(1,2,3,4-tetrahydronaphthalen-1-yl)nicotinamide	4.19		

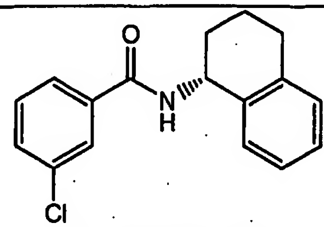
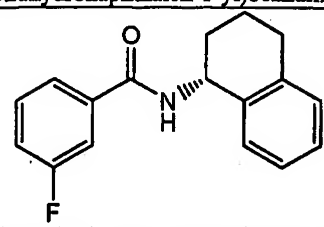
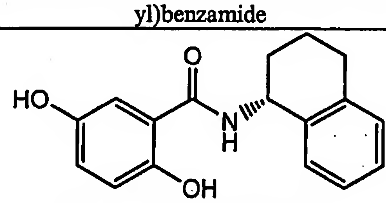
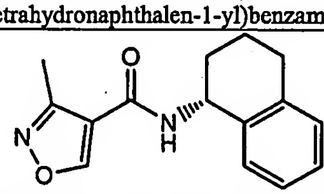
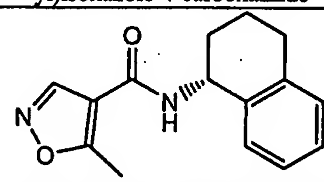
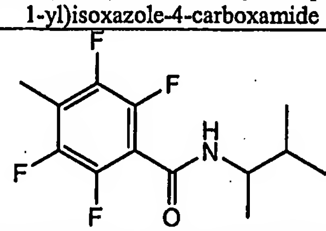
Table 5 - Sweet EnhancerAmides				
Compound No.	Compound	Sweet EC_{50} uM	Umami EC_{50} uM	Umami EC_{50} ratio
E22	 (R)-3-chloro-N-(1,2,3,4-tetrahydronaphthalen-1-yl)benzamide	4.52		
E23	 (R)-3-fluoro-N-(1,2,3,4-tetrahydronaphthalen-1-yl)benzamide	4.86		
E24	 (R)-2,5-dihydroxy-N-(1,2,3,4-tetrahydronaphthalen-1-yl)benzamide	6.04		
E25	 (R)-3-methyl-N-(1,2,3,4-tetrahydronaphthalen-1-yl)isoxazole-4-carboxamide	7.79		
E26	 (R)-5-methyl-N-(1,2,3,4-tetrahydronaphthalen-1-yl)isoxazole-4-carboxamide	8.09		
E27	 2,3,5,6-tetrafluoro-4-methyl-N-(3-methylbutan-2-yl)benzamide	0.14		

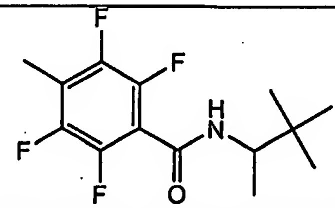
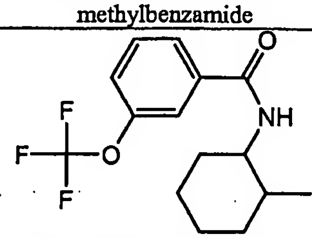
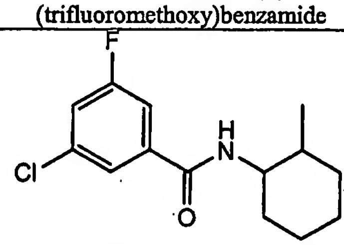
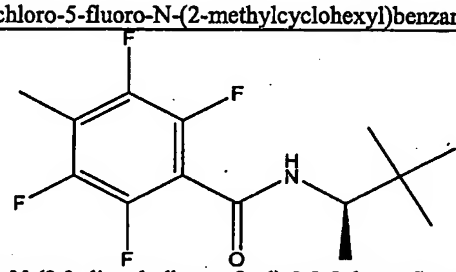
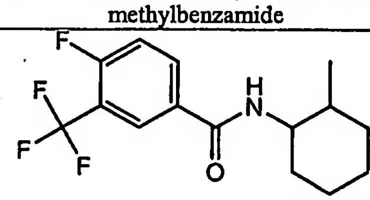
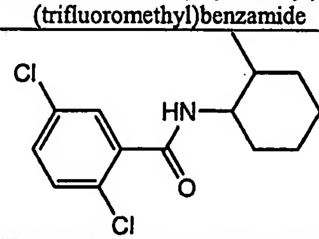
Table 5 - Sweet EnhancerAmides				
Compound No.	Compound	Sweet EC_{50} uM	Umami EC_{50} uM	Umami EC_{50} ratio
E28	 N-(3,3-dimethylbutan-2-yl)-2,3,5,6-tetrafluoro-4-methylbenzamide	0.21		
E29	 N-(2-methylcyclohexyl)-3-(trifluoromethoxy)benzamide	0.42		
E30	 3-chloro-5-fluoro-N-(2-methylcyclohexyl)benzamide	0.45		
E31	 (R)-N-(3,3-dimethylbutan-2-yl)-2,3,5,6-tetrafluoro-4-methylbenzamide	0.49		
E32	 4-fluoro-N-(2-methylcyclohexyl)-3-(trifluoromethyl)benzamide	0.51		
E33	 2,5-dichloro-N-(2-methylcyclohexyl)benzamide	0.63		

Table 5 - Sweet EnhancerAmides

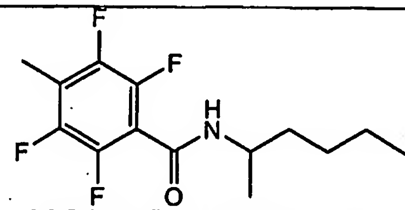
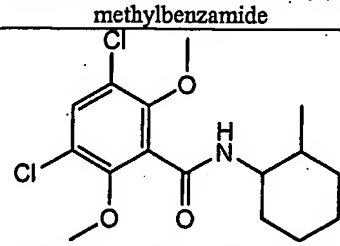
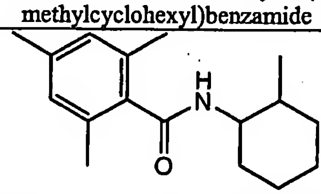
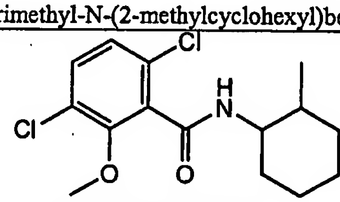
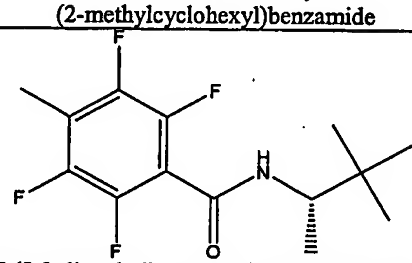
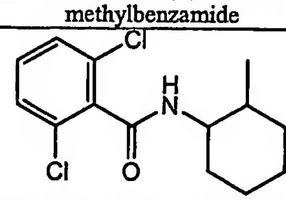
Compound No.	Compound	Sweet EC_{50} uM	Umami EC_{50} uM	Umami EC_{50} ratio
E34	 2,3,5,6-tetrafluoro-N-(hexan-2-yl)-4-methylbenzamide	0.71		
E35	 3,5-dichloro-2,6-dimethoxy-N-(2-methylcyclohexyl)benzamide	0.71		
E36	 2,4,6-trimethyl-N-(2-methylcyclohexyl)benzamide	0.72		
E37	 3,6-dichloro-2-methoxy-N-(2-methylcyclohexyl)benzamide	0.77		
E38	 (S)-N-(3,3-dimethylbutan-2-yl)-2,3,5,6-tetrafluoro-4-methylbenzamide	0.9		
E39	 2,6-dichloro-N-(2-methylcyclohexyl)benzamide	0.91		

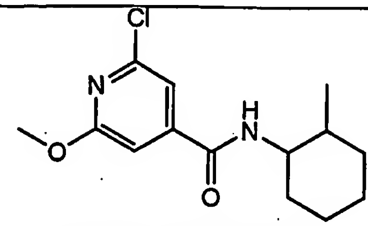
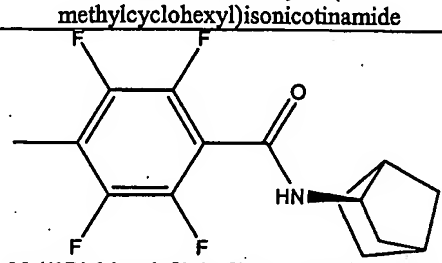
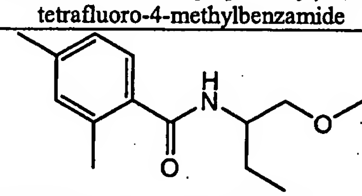
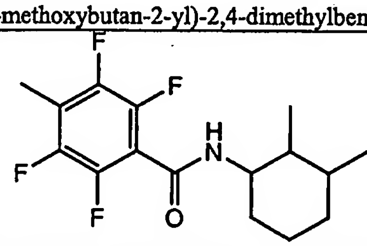
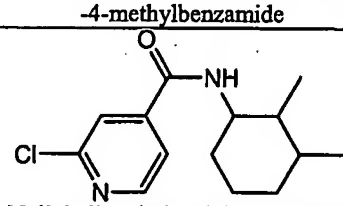
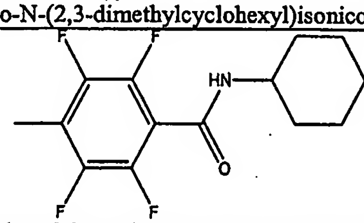
Table 5 - Sweet EnhancerAmides				
Compound No.	Compound	Sweet EC_{50} uM	Umami EC_{50} uM	Umami EC_{50} ratio
E40	 2-chloro-6-methoxy-N-(2-methylcyclohexyl)isonicotinamide	0.95		9.77
E41	 N-((2R)-bicyclo[2.2.1]heptan-2-yl)-2,3,5,6-tetrafluoro-4-methylbenzamide	1.02		
E42	 N-(1-methoxybutan-2-yl)-2,4-dimethylbenzamide	1.06		
E43	 N-(2,3-dimethylcyclohexyl)-2,3,5,6-tetrafluoro -4-methylbenzamide	1.08		
E44	 2-chloro-N-(2,3-dimethylcyclohexyl)isonicotinamide	1.08		
E45	 N-cyclohexyl-2,3,5,6-tetrafluoro-4-methylbenzamide	1.13		

Table 5 - Sweet EnhancerAmides

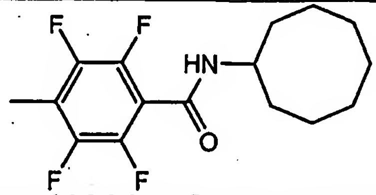
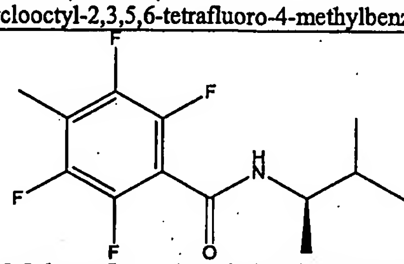
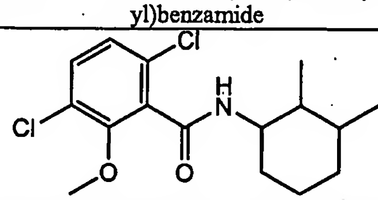
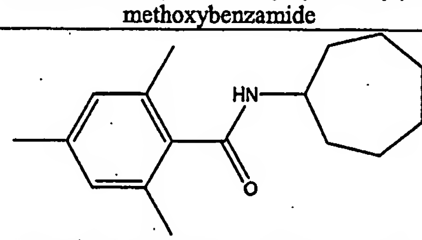
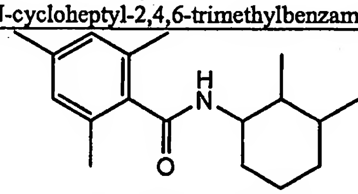
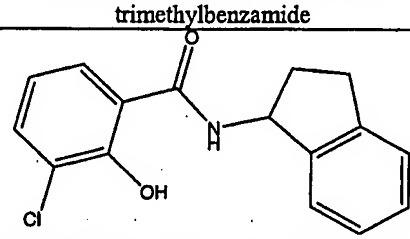
Compound No.	Compound	Sweet EC_{50} uM	Umami EC_{50} uM	Umami EC_{50} ratio
E46	 N-cyclooctyl-2,3,5,6-tetrafluoro-4-methylbenzamide	1.25		
E47	 (R)-2,3,5,6-tetrafluoro-4-methyl-N-(3-methylbutan-2-yl)benzamide	1.25		
E48	 3,6-dichloro-N-(2,3-dimethylcyclohexyl)-2-methoxybenzamide	1.29		
E49	 N-cycloheptyl-2,4,6-trimethylbenzamide	1.39		
E50	 N-(2,3-dimethylcyclohexyl)-2,4,6-trimethylbenzamide	1.41		
E51	 3-chloro-N-(2,3-dihydro-1H-inden-1-yl)-2-hydroxybenzamide	1.49		

Table 5 - Sweet EnhancerAmides

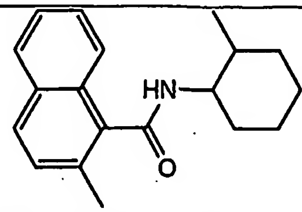
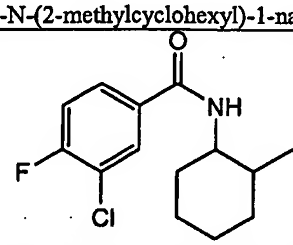
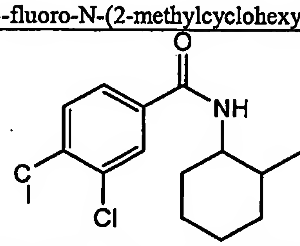
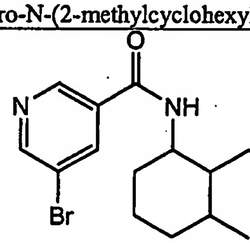
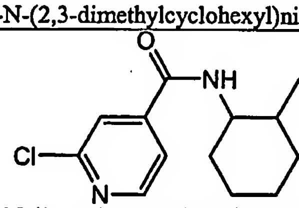
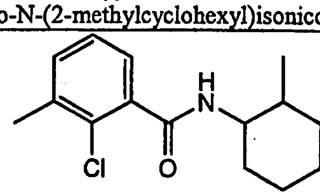
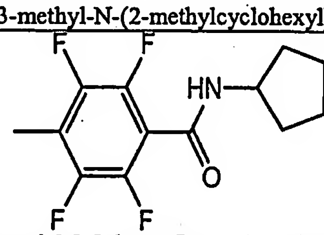
Compound No.	Compound	Sweet EC_{50} uM	Umami EC_{50} uM	Umami EC_{50} ratio
E52	 2-methyl-N-(2-methylcyclohexyl)-1-naphthamide	1.52		
E53	 3-chloro-4-fluoro-N-(2-methylcyclohexyl)benzamide	1.7		
E54	 3,4-dichloro-N-(2-methylcyclohexyl)benzamide	1.83		10.66
E55	 5-bromo-N-(2,3-dimethylcyclohexyl)nicotinamide	1.89		
E56	 2-chloro-N-(2-methylcyclohexyl)isonicotinamide	1.92		2.08
E57	 2-chloro-3-methyl-N-(2-methylcyclohexyl)benzamide	1.95		
E58	 N-cyclopentyl-2,3,5,6-tetrafluoro-4-methylbenzamide	2.23		

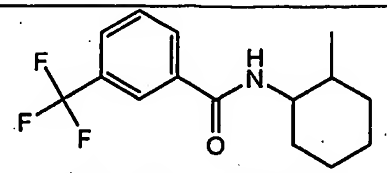
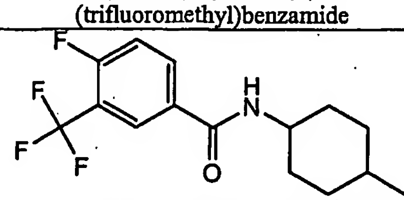
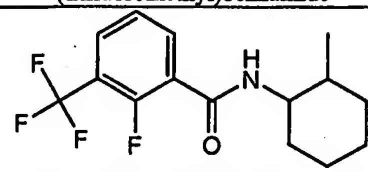
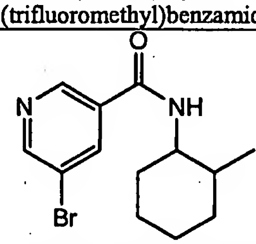
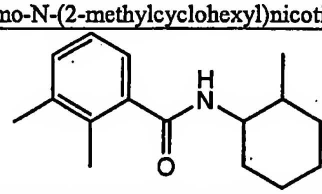
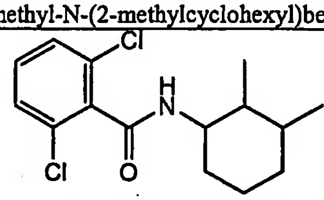
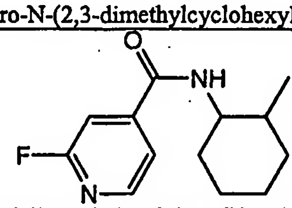
Table 5 - Sweet EnhancerAmides				
Compound No.	Compound	Sweet EC50 uM	Umami EC50 uM	Umami EC50 ratio
E59	 N-(2-methylcyclohexyl)-3-(trifluoromethyl)benzamide	2.34		2.07
E60	 4-fluoro-N-(4-methylcyclohexyl)-3-(trifluoromethyl)benzamide	2.37		
E61	 2-fluoro-N-(2-methylcyclohexyl)-3-(trifluoromethyl)benzamide	2.4		
E62	 5-bromo-N-(2-methylcyclohexyl)nicotinamide	2.42		
E63	 2,3-dimethyl-N-(2-methylcyclohexyl)benzamide	2.6		
E64	 2,6-dichloro-N-(2,3-dimethylcyclohexyl)benzamide	2.77		
E65	 2-fluoro-N-(2-methylcyclohexyl)isonicotinamide	2.83		

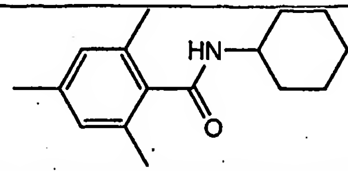
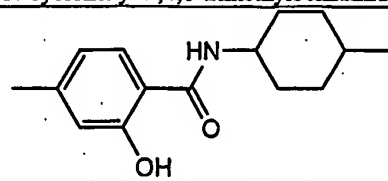
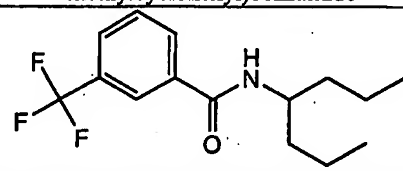
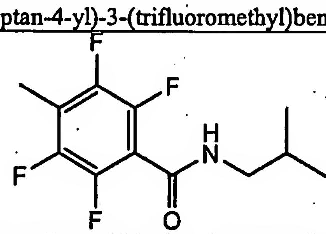
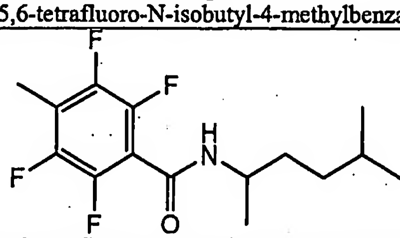
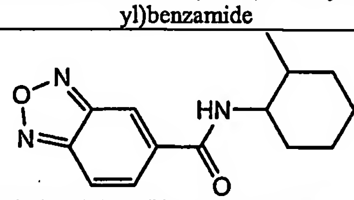
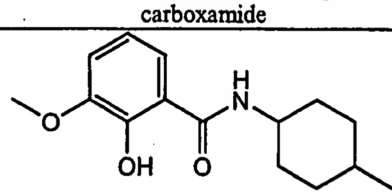
Table 5 - Sweet EnhancerAmides				
Compound No.	Compound	Sweet EC_{50} uM	Umami EC_{50} uM	Umami EC_{50} ratio
E66	 N-cyclohexyl-2,4,6-trimethylbenzamide	2.86		
E67	 2-hydroxy-4-methyl-N-(4-methylcyclohexyl)benzamide	2.98		
E68	 N-(heptan-4-yl)-3-(trifluoromethyl)benzamide	3.03	0.33	
E69	 2,3,5,6-tetrafluoro-N-isobutyl-4-methylbenzamide	3.19		
E70	 2,3,5,6-tetrafluoro-4-methyl-N-(5-methylhexan-2-yl)benzamide	3.2		
E71	 N-(2-methylcyclohexyl)benzo[c][1,2,5]oxadiazole-5-carboxamide	3.33		
E72	 2-hydroxy-3-methoxy-N-(4-methylcyclohexyl)benzamide	3.35		

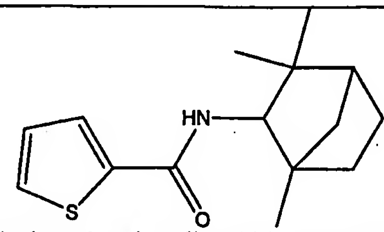
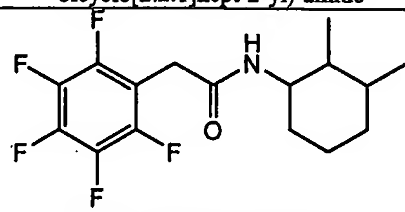
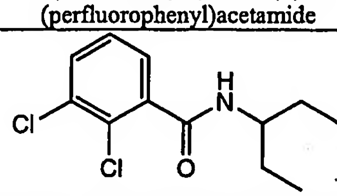
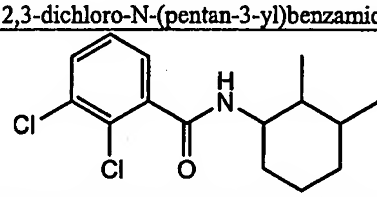
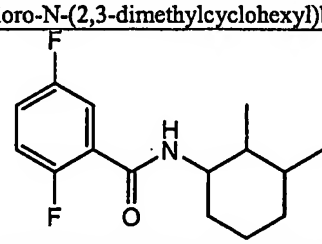
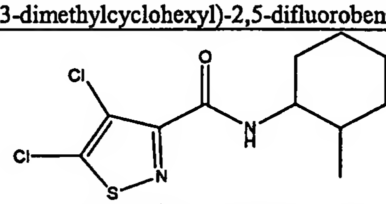
Table 5 - Sweet EnhancerAmides				
Compound No.	Compound	Sweet EC_{50} uM	Umami EC_{50} uM	Umami EC_{50} ratio
E73	 Thiophene-2-carboxylic acid (1,3,3-trimethyl-bicyclo[2.2.1]hept-2-yl)-amide	3.36		
E74	 N-(2,3-dimethylcyclohexyl)-2-(perfluorophenyl)acetamide	3.62		
E75	 2,3-dichloro-N-(pentan-3-yl)benzamide	3.78		
E76	 2,3-dichloro-N-(2,3-dimethylcyclohexyl)benzamide	3.99		
E77	 N-(2,3-dimethylcyclohexyl)-2,5-difluorobenzamide	4.11		
E78	 4,5-Dichloro-isothiazole-3-carboxylic acid (2-methyl-cyclohexyl)-amide	4.24	8.51	

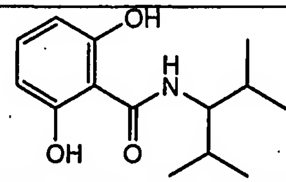
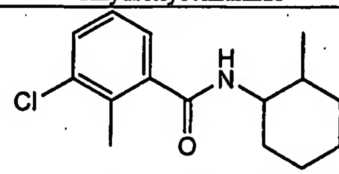
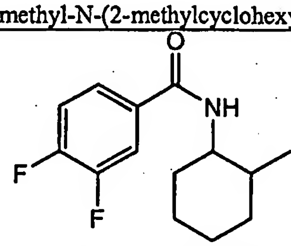
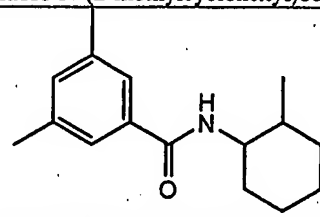
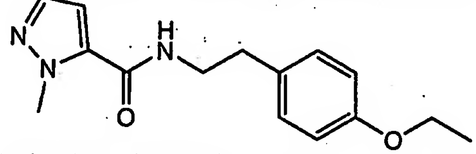
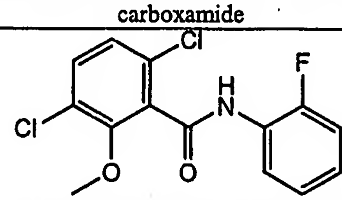
Table 5 - Sweet EnhancerAmides				
Compound No.	**Compound**	**Sweet EC_{50} uM**	**Umami EC_{50} uM**	**Umami EC_{50} ratio**
E79	 N-(2,4-dimethylpentan-3-yl)-2,6-dihydroxybenzamide	4.28		
E80	 3-chloro-2-methyl-N-(2-methylcyclohexyl)benzamide	4.29		
E81	 3,4-difluoro-N-(2-methylcyclohexyl)benzamide	4.37		6.98
E82	 3,5-dimethyl-N-(2-methylcyclohexyl)benzamide	4.48		
E83	 N-(4-ethoxyphenethyl)-1-methyl-1H-pyrazole-5-carboxamide	4.68		
E84	 3,6-dichloro-N-(2-fluorophenyl)-2-methoxybenzamide	0.83		16.51

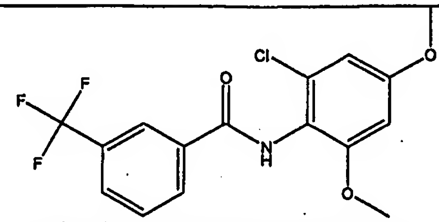
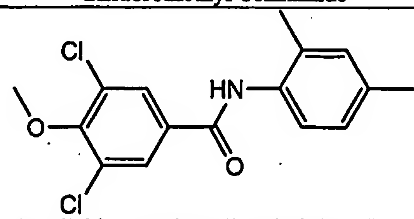
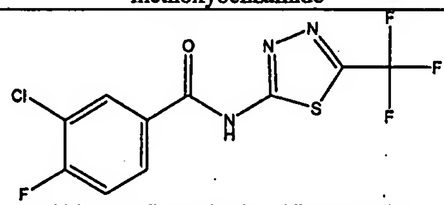
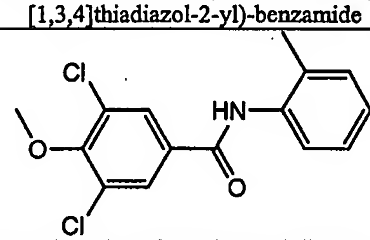
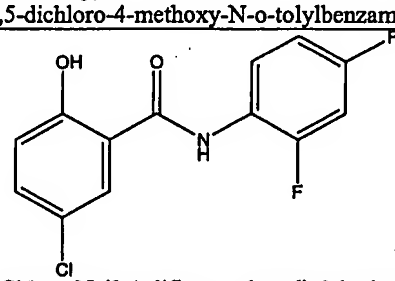
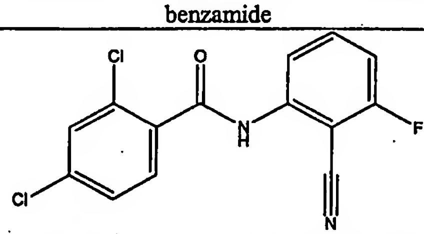
Table 5 - Sweet EnhancerAmides				
Compound No.	Compound	Sweet EC_{50} uM	Umami EC_{50} uM	Umami EC_{50} ratio
E85	 N-(2-Chloro-4,6-dimethoxy-phenyl)-3-trifluoromethyl-benzamide	1.42		
E86	 3,5-dichloro-N-(2,4-dimethylphenyl)-4-methoxybenzamide	1.48		
E87	 3-Chloro-4-fluoro-N-(5-trifluoromethyl-[1,3,4]thiadiazol-2-yl)-benzamide	1.55		
E88	 3,5-dichloro-4-methoxy-N-o-tolylbenzamide	1.84		
E89	 5-Chloro-N-(2,4-difluoro-phenyl)-2-hydroxy-benzamide	2.56		
E90	 2,4-Dichloro-N-(2-cyano-3-fluoro-phenyl)-benzamide	2.71		

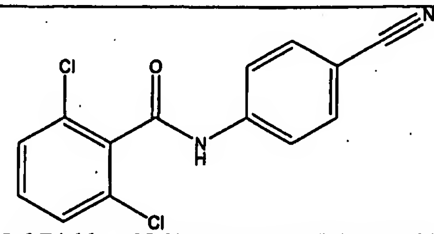
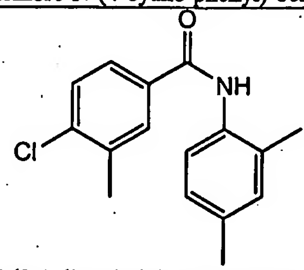
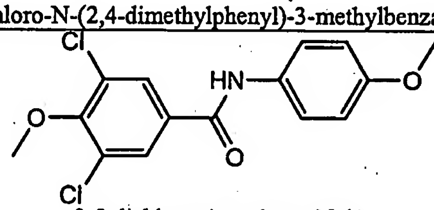
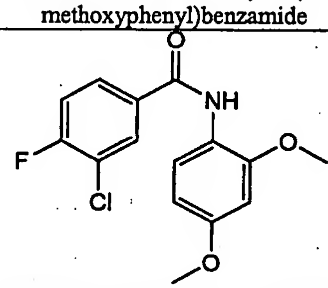
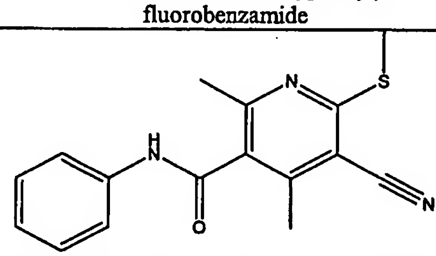
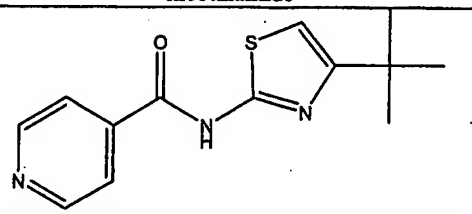
Table 5 - Sweet EnhancerAmides				
Compound No.	Compound	Sweet EC_{50} uM	Umami EC_{50} uM	Umami EC_{50} ratio
E91	 2,6-Dichloro-N-(4-cyano-phenyl)-benzamide	2.74		
E92	 4-chloro-N-(2,4-dimethylphenyl)-3-methylbenzamide	2.74		
E93	 3,5-dichloro-4-methoxy-N-(4-methoxyphenyl)benzamide	3.24		
E94	 3-chloro-N-(2,4-dimethoxyphenyl)-4-fluorobenzamide	3.56		
E95	 5-Cyano-2,4-dimethyl-6-methylsulfanyl-N-phenyl-nicotinamide	3.58		
E96	 N-(4-tert-Butyl-thiazol-2-yl)-isonicotinamide	3.73		

Table 5 - Sweet EnhancerAmides

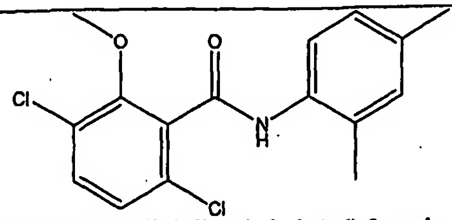
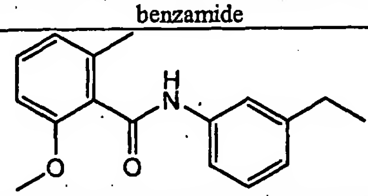
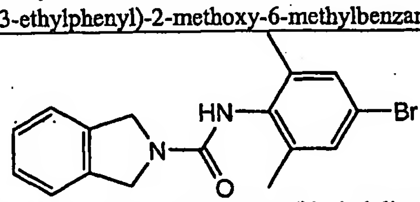
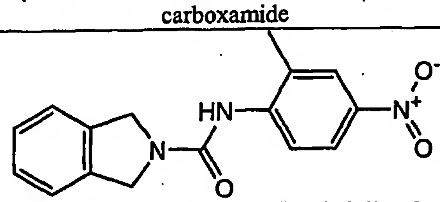
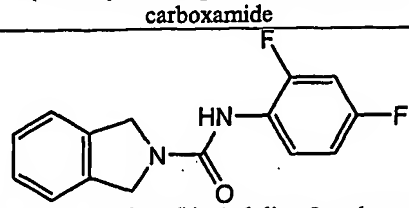
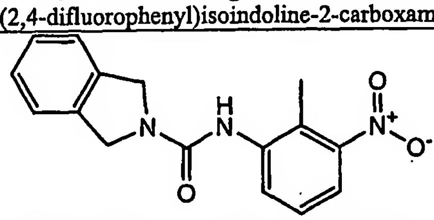
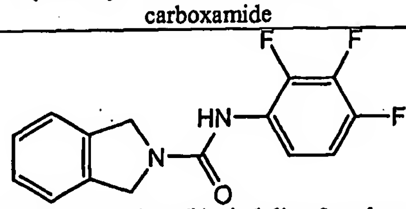
Compound No.	Compound	Sweet EC_{50} uM	Umami EC_{50} uM	Umami EC_{50} ratio
E97	 3,6-Dichloro-N-(2,4-dimethyl-phenyl)-2-methoxy-benzamide	4.25		
E98	 N-(3-ethylphenyl)-2-methoxy-6-methylbenzamide	4.63		
E99	 N-(4-bromo-2,6-dimethylphenyl)isoindoline-2-carboxamide	0.93		
E100	 N-(2-methyl-4-nitrophenyl)isoindoline-2-carboxamide	1.3		
E101	 N-(2,4-difluorophenyl)isoindoline-2-carboxamide	1.37		
E102	 N-(2-methyl-3-nitrophenyl)isoindoline-2-carboxamide	2.01		
E103	 N-(2,3,4-trifluorophenyl)isoindoline-2-carboxamide	2.58		

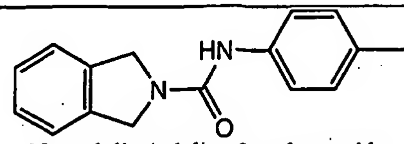
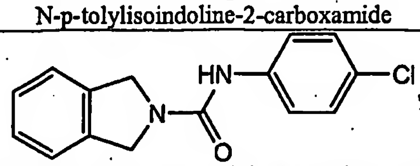
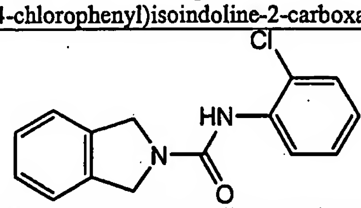
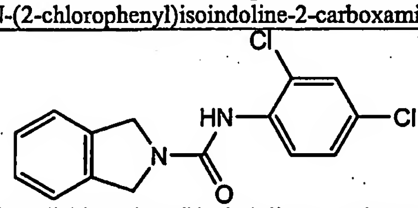
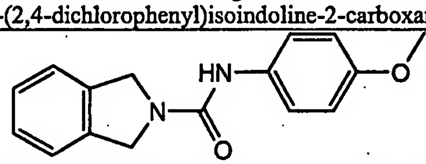
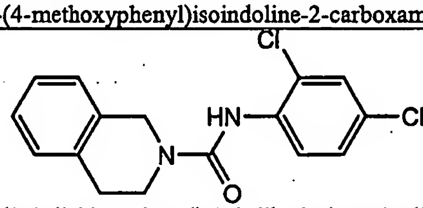
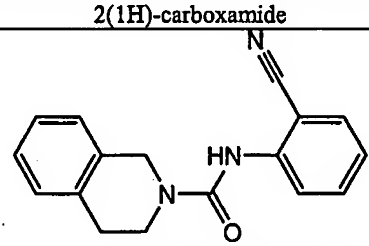
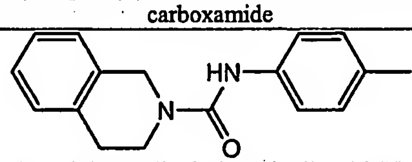
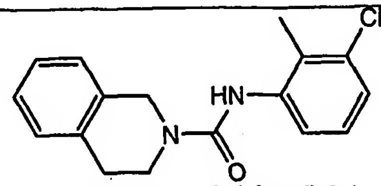
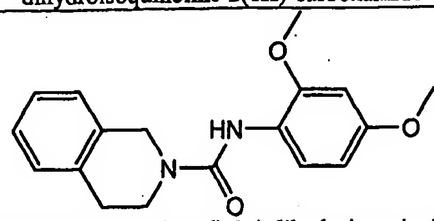
Table 5 - Sweet EnhancerAmides				
Compound No.	**Compound**	**Sweet EC50 uM**	**Umami EC50 uM**	**Umami EC50 ratio**
E104	 N-p-tolylisoindoline-2-carboxamide	3.05		
E105	 N-(4-chlorophenyl)isoindoline-2-carboxamide	3.4		
E106	 N-(2-chlorophenyl)isoindoline-2-carboxamide	3.85		
E107	 N-(2,4-dichlorophenyl)isoindoline-2-carboxamide	4.15		
E108	 N-(4-methoxyphenyl)isoindoline-2-carboxamide	4.99		
E109	 N-(2,4-dichlorophenyl)-3,4-dihydroisoquinoline-2(1H)-carboxamide	2.34		
E110	 N-(2-cyanophenyl)-3,4-dihydroisoquinoline-2(1H)-carboxamide	2.5		
E111	 N-p-tolyl-3,4-dihydroisoquinoline-2(1H)-carboxamide	4.27		

Table 5 - Sweet EnhancerAmides				
Compound No.	Compound	Sweet EC_{50} uM	Umami EC_{50} uM	Umami EC_{50} ratio
E112	 N-(3-chloro-2-methylphenyl)-3,4-dihydroisoquinoline-2(1H)-carboxamide	4.33		
E113	 N-(2,4-dimethoxyphenyl)-3,4-dihydroisoquinoline-2(1H)-carboxamide	4.44		

Also, as supported by experimental data in the examples, it has been shown that cell lines which co-express T1R1/T1R3 or T1R2/T1R3 respectively respond to umami or sweet taste stimuli and a quantitative dose-responsive manner which further supports a conclusion that specific binding to the T1R1/T1R3 and T1R2/T1R3 receptor can be used to define receptor agonists and antagonists, e.g., MSG substitutes, umami blockers, novel artificial and natural sweeteners, and sweet blockers.

Also, as supported by data in experimental examples, it has been shown that the sweet taste blocker lactisole inhibits both the T1R2/T1R3 sweet receptor and the T1R1/T1R3 umami taste receptor. Compounds are provided herein that enhance, mimic, modulate or block sweet or umami taste. The fact that lactisole inhibits both the T1R1/T1R3 and T1R2/T1R3 receptors suggests that these receptors may share a common subunit which is bound by lactisole and potentially other taste modulators. Therefore, this shows that some compounds which enhance, mimic, modulate or block sweet taste can have a similar effect on umami taste or vice versa.

Further, as supported by data in experimental examples, it has been demonstrated that cell lines which stably co-express T1Rs, i.e. T1R1/T1R3 or T1R2/T1R3, when assayed by automated fluorescence imaging very effectively respond to various sweet and umami taste stimuli, i.e. at magnitudes substantially greater than transiently transfected cells. Thus, these cell lines are especially well suited for use in high throughput screening assays for identifying compounds that modulate, block, mimic or enhance sweet or umami taste. However, the invention also encompasses assays that utilize cells that transiently express a T1R or combination thereof.

Moreover, while the application contains data demonstrating that some T1Rs act in combination, particularly T1R1/T1R3 and T1R2/T1R3, and that such receptor combinations may be used in assays, preferably high throughput assays, it should be noted that the subject invention also encompasses assays that utilize T1R1, T1R2 and T1R3 alone or in combination with other proteins, e.g., other GPCRs.

There are differences in human and rodent sweet taste in terms of the ligand specificity, G protein coupling efficiency, as well as sensitivity to inhibitors. The species differences in T1R ligand specificity can be utilized to demonstrate that the sweet taste receptor indeed functions as a heteromeric complex, and that there is more than one ligand binding site on the receptor. Furthermore, a functional link between

the sweet and umami receptors mediated by T1R3 has been shown (Example 16).

Both human and rat sweet receptors can efficiently couple to a chimeric Gα15 with the C-terminal tail sequence from $G_{\alpha il}$ ($G_{\alpha 15/il}$). For example, human but not rat T1R2/T1R3 selectively responds to a group of sweeteners, including aspartame, neotame, and cyclamate. This is consistent with taste physiology data. These differences in agonist specificity can be utilized to map their binding sites on the receptor. A chimeric T1R can be generated between human and rat genes, with a junction immediately before the transmembrane domain. Each T1R chimera therefore consists of two halves, the N-terminal extracellular domain, and the C-terminal transmembrane and intracellular domain, from different species. For example, a chimeric T1R2, termed T1R2-R, has a sequence from the N-terminus of human T1R2 fused to rat T1R2 C-terminal sequence. Responses to these chimeras can then be tested (Figure 22).

Novel compounds and novel flavor, tastants, and sweet enhancers were discovered in the chemistry series of amide derivatives. The amide compounds also comprise certain sub-classes of amide derivatives or classes of derivatives related to amides, such as for example ureas, urethanes, oxalamides, acrylamides, and the like. These compounds, when used together with sucrose or alone, increase a response in vitro and concomitant increase in sweet perception in human tasting. These compounds enhance other natural and synthetic sweet tastants. Examples of these compounds are listed in Table 5.

In one embodiment, the invention provides novel compounds, flavorants, tastants, flavor enhancers, taste enhancers, flavor modifying compounds, and/or compositions containing them.

In a more specific embodiment, the invention provides novel sweet flavorants, sweet tastants, sweet taste enhancers, and sweet taste modifiers and compositions containing them.

More particularly, in another embodiment, the invention is directed to compounds that modulate, induce, enhance, or inhibit natural or synthetic sweet tastants, *e.g.*, naturally occurring and synthetic sweeteners.

In another embodiment, the invention provides compositions, preferably compositions suitable for human or animal consumption, containing at least one compound of the invention. These compositions include foods, beverages and

medicinals, and food additives which when added to foods, beverages or medicinals modulate the flavor or taste thereof, particularly by enhancing the sweet taste thereof.

Another embodiment of the invention is directed to use of a compound of the invention to modulate the sweet taste of a desired food, beverage or medicinal, which composition may comprise one or more other compounds that elicit a sweet taste.These compounds, when they were used together with naturally occurring and synthetic sweeteners, not only increased a response *in vitro* but also intensified the sweet and other flavor or taste perceptions in human tasting. These specific compounds, when they were used together with sweet tastants, such as naturally occurring and synthetic sweeteners, not only increased the T1R2/T1R3 response *in vitro* but also intensified the sweet taste and other flavor or taste perceptions in human tasting.

Novel compounds and novel flavor, tastant, and umami enhancers and tastants such as amides, ureas, amino-amides, amido-amides, and β-lactams are also disclosed herein. These compounds, when used together with MSG or alone, increase a response *in vitro* and the umami perception in human tasting. These compounds also enhance other natural and synthetic umami tastants. Examples of these compounds are listed in Tables 1-4.

In one embodiment, the invention provides novel compounds, flavorants, tastants, flavor enhancers, taste enhancers, flavor modifying compounds, and/or compositions containing them.

In a more specific embodiment, the invention provides novel umami flavorants, umami tastants, umami taste enhancers, and umami taste modifiers and compositions containing them.

More particularly, in another embodiment, the invention is directed to compounds that modulate (induce, enhance or inhibit) natural or synthetic umami tastants, *e.g.*, monosodium glutamate (MSG).

In another embodiment, the invention provides compositions, preferably compositions suitable for human or animal consumption, containing at least one compound of the invention. These compositions include foods, beverages and medicinals, and food additives which when added to foods, beverages or medicinals modulate the flavor or taste thereof, particularly by enhancing the umami taste thereof.

Another embodiment of the invention is directed to use of a compound of the invention to modulate the umami taste of a desired food, beverage or medicinal, which

composition may comprise one or more other compounds that elicit a umami taste, *e.g.*, MSG. These compounds, when they were used together with MSG, not only increased a response *in vitro* but also intensified the umami and other flavor or taste perceptions in human tasting. These specific compounds, when they were used together with umami tastants, such as MSG, not only increased the T1R1/T1R3 response *in vitro* but also intensified the umami taste and other flavor or taste perceptions in human tasting. Some of the compounds, when they were tasted alone, elicited human perception of umami.

Compounds defined by specific binding to specific receptors using the present T1R assays can be used to modulate the taste of foods and beverages. Suitable assays described in further detail infra include by way of example whole-cell assays and biochemical assays, including direct-binding assays using one of a combination of different T1R receptors, chimeras or fragments thereof, especially fragments containing N-terminal ligand-binding domains. Examples of assays appropriate for use in the invention are described in greater detail infra and are known in the GPCR field.

Assays can be designed that quantitate the binding of different compounds or mixtures of compounds to T1R taste receptors or T1R taste receptor combinations or T1R receptors expressed in combination with other heterologous (non-T1R) proteins, e.g. other GPCRs, or that quantitate the activation of cells that express T1R taste receptors. This can be effected by stably or transiently expressing taste receptors in heterologous cells such as HEK-293, CHO and COS cells. Thus, this physico-chemical characteristic of the compounds is used to define a genus of compound that share this characteristic.

The assays will preferably use cells that also express (preferably stably) a G protein such as Gα15 or Gα16 or other promiscuous G proteins or G protein variants, or an endogenous G protein. In addition, G_β and G_γ proteins may also be expressed therein.

The effect of a compound on sweet or umami taste using cells or compositions that express or contain the above-identified receptors or receptor combinations may be determined by various means including the use of calcium-sensitive dyes, voltage-sensitive dyes, cAMP assays, direct binding assays using fluorescently labeled ligands

or radioactive ligands such as ^{3}H-glutamate, or transcriptional assays (using a suitable reporter such as luciferase or beta-lactamase).

Assays that may be utilized with one or more T1Rs according to the invention include by way of example, assays that utilize a genetic selection for living cells; assays that utilize whole cells or membrane fragments or purified T1R proteins; assays that utilize second messengers such as cAMP and IP3, assays that detect the translocation of arrestin to the cell surface, assays that detect the loss of receptor expression on the cell surface (internalization) by tested ligands, direct ligand-binding assays, competitive-binding assays with inhibitors, assays using in vitro translated protein, assays that detect conformational changes upon the binding of a ligand (e.g., as evidenced by proteolysis, fluorescence, or NMR), behavioral assays that utilize transgenic non-human animals that express a T1R or T1R combination, such as flies, worms, or mice, assays that utilize cells infected with recombinant viruses that contain T1R genes.

Also within the scope of the invention are structure-based analyses wherein the X-ray crystal structure of a T1R or T1R fragment (or combination of T1Rs, or a combination of a T1R with another protein) is determined and utilized to predict by molecular modeling techniques compounds that will bind to and/or enhance, mimic, block or modulate the particular T1R receptor or receptor combination. More particularly, the invention embraces the determination of the crystal structure of T1R1/T1R3 (preferably hT1R1/hT1R3) and/or T1R2/T1R3 (preferably hT1R2/hT1R3) and the use of such crystal structures in structure-based design methods to identify molecules that modulate T1R receptor activity.

The invention especially includes biochemical assays conducted using cells, e.g., mammalian, yeast, insect or other heterologous cells that express one or more full length T1R receptors or fragments, preferably N-terminal domains of T1R1, T1R2 and/or T1R3. The effect of a compound in such assays can be determined using competitive binding assays, e.g., using radioactive glutamate or IMP, fluorescence (e.g., fluorescence polarization, FRET), or GTPγ ^{35}S binding assays. As noted, in a preferred embodiment, such assays will utilize cell lines that stably co-express T1R1/T1R3 or T1R2/T1R3 and a suitable G protein, such as $G_{\alpha 15}$. Other appropriate G proteins include the chimeric and variant G proteins disclosed in U.S. Application Serial No. 09/984,292 and 60/243,770, incorporated by reference in their entirety herein.

Still further, altered receptors can be constructed and expressed having improved properties, e.g., enhanced surface expression or G-protein coupling. These T1R variants can be incorporated into cell-based and biochemical assays.

It is envisioned that the present discoveries relating to human T1Rs will extend to other species, e.g., rodents, pigs, monkeys, dogs and cats, and perhaps even non-mammals such as fish. In this regard, several fish T1R fragments are identified infra in Example 1. Therefore, the subject invention has application in screening for compounds for use in animal feed formulations.

The invention further includes that utilize different allelic variants of various T1Rs and combinations thereof, thereby enabling the identification of compounds that elicit specific taste sensation in individuals that express those allelic variants or compounds that elicit specific taste sensations in all individuals. Such compounds can be used to make foods more generally palatable.

T1R encoding nucleic acids also provide valuable probes for the identification of taste cells, as the nucleic acids are specifically expressed in taste cells. For example, probes for T1R polypeptides and proteins can be used to identify taste cells present in foliate, circumvallate, and fungiform papillae, as well as taste cells present in the geschmackstreifen, oral cavity, gastrointestinal epithelium, and epiglottis. In particular, methods of detecting T1Rs can be used to identify taste cells sensitive to sweet and/or umami taste stimuli or other taste stimuli representing other taste modalities. For example, cells stably or transiently expressing T1R2 and/or T1R3 would be predicted from the work herein to be responsive to sweet taste stimuli. Similarly, cells expressing T1R1 and/or T1R3 would be predicted to be responsive to umami taste stimuli. The nucleic acids encoding the T1R proteins and polypeptides of the invention can be isolated from a variety of sources, genetically engineered, amplified, synthesized, and/or expressed recombinantly according to the methods disclosed in WO 00/035374, which is herein incorporated by reference in its entirety. A listing of T1Rs that may be expressed according to the invention are provided in the Examples. However, it should be emphasized that the invention embraces the expression and use of other specific T1Rs or fragments, variants, or chimeras constructed based on such T1R sequences, and particularly T1Rs of other species.

As disclosed, an important aspect of the invention is the plurality of methods of screening for modulators, *e.g.*, activators, inhibitors, stimulators, enhancers, agonists,

and antagonists, of these taste-cell-specific GPCRs. Such modulators of taste transduction are useful for the modulation of taste signaling pathways. These methods of screening can be used to identify high affinity agonists and antagonists of taste cell activity. These modulatory compounds can then be used in the food industry to customize taste, *e.g.*, to modulate the sweet and/or umami tastes of foods.

This invention rectifies the previous lack of understanding relating to sweet and umami taste as it identifies specific T1Rs and T1R receptor combinations that mediate sweet and umami taste sensation. Therefore, in general, this application relates to the inventors' discoveries relating to the T1R class of taste-specific G-protein-coupled receptors and their specific function in taste perception and the relationship of these discoveries to a better understanding of the molecular basis of taste.

The molecular basis of sweet taste and umami taste – the savor of monosodium glutamate – is enigmatic. Recently, a three-member class of taste-specific G-protein-coupled receptors, termed T1Rs, was identified. Overlapping T1R expression patterns and the demonstration that the structurally related $GABA_B$ receptor is heterodimeric suggest that the T1Rs function as heterodimeric taste receptors. In the examples infra, the present inventors describe the functional co-expression of human T1R1, T1R2, and T1R3 in heterologous cells; cells co-expressing T1R1 and T1R3 are activated by umami taste stimuli; cells co-expressing T1R2 and T1R3 are activated by sweet taste stimuli. T1R1/T1R3 and T1R2/T1R3 activity correlated with psychophysical detection thresholds. In addition, the 5'-ribonucleotide IMP was found to enhance the T1R1/T1R3 response to glutamate, a synergism characteristic of umami taste. These findings demonstrate that specific T1Rs and particularly different combinations of the T1Rs function as sweet and umami taste receptors.

Human perception of bitter, sweet, and umami is thought to be mediated by G-protein-coupled receptors (Lindemann, B., *Physiol. Res.* 76:718-66 (1996)). Recently, evaluation of the human genome revealed the T2R class of bitter taste receptors (Adler et al., *Cell* 100:613-702 (2000); Chandrasgekar et al., *Cell* 100:703-11 (2000); Matsunami et al., *Nature* 404: 601-604 (2000)) but the receptors for sweet and umami taste have not been identified. Recently, another class of candidate taste receptors, the T1Rs, was identified. The T1Rs were first identified by large-scale sequencing of a subtracted cDNA library derived from rat taste tissue, which identified T1R1, and subsequently by T1R1-based degenerate PCR, which led to the identification of T1R2

(Hoon et al., *Cell* 96:541-551 (1999)). Recently, the present inventors and others identified a third and possibly final member of the T1R family, T1R3, in the human genome databank (Kitagawa et al., *Biochem Biophys. Res Commun.* 283(1): 236-42 (2001); Max et al., *Nat. Genet.* 28(1): 58-63 (2001); Sainz et al., *J. Neurochem.* 77(3): 896-903 (2001); Montmayeur et al., *Nat. Neurosci.* 4, 492-8. (2001)). Tellingly, mouse T1R3 maps to a genomic interval containing *Sac*, a locus that influences sweet taste in the mouse (Fuller et al., *J. Hered.* 65:33-6 (1974); Li et al., *Mamm. Genome* 12:13-16 (2001)). Therefore, T1R3 was predicted to function as a sweet taste receptor. Recent high-resolution genetic mapping studies have strengthened the connection between mouse T1R3 and *Sac* (Fuller T.C., *J. Hered.* 65(1): 33-36 (1974); Li et al., *Mammal. Genome* 12(1): 13-16 (2001)).

Interestingly, all C-family receptors that have been functionally expressed thus far – metabotropic glutamate receptors, the $GABA_B$ receptor, the calcium-sensing receptor (Conigrave, A. D., Quinn, S. J. & Brown, E. M., Proc Natl Acad Sci *U S A* 97, 4814-9. (2000)), and a fish olfactory receptor (Speca, D. J. et al., *Neuron* 23, 487-98. (1999)) – have been shown to be activated by amino acids. This common feature raises the possibility that the T1Rs recognize amino acids, and that the T1Rs may be involved in the detection of glutamate in addition to sweet-tasting amino acids. Alternatively, a transcriptional variant of the mGluR4 metabotropic glutamate receptor has been proposed to be the umami taste receptor because of its selective expression in rat taste tissue, and the similarity of the receptor-activation threshold to the glutamate psychophysical detection threshold (Chaudhari et al., *Nat. Neurosci.* 3:113-119 (2000)). This hypothesis is difficult to reconcile with the exceedingly low expression level of the mGluR4 variant in taste tissue, and the more or less unaltered glutamate taste of mGluR4 knockout mice (Chaudhari and Roper, *Ann. N.Y. Acad. Sci.* 855:398-406 (1998)). Furthermore, the taste variant is structurally implausible, lacking not only the majority of the residues that form the glutamate-binding pocket of the wild-type receptor, but also approximately half of the globular N-terminal glutamate-binding domain (Kunishima et al., *Nature* 407:971-7 (2000)).

Comparative analysis of T1R expression patterns in rodents has demonstrated that T1R2 and possibly T1R1 are each coexpressed with T1R3 (Hoon et al., *Cell* 96:541-51 (1999); Kitagawa et al., *Biochem Biophy. Res. Commun.* 283:236-242 (2001); Max et al., *Nat. Genet.* 28:58-63 (2001); Montmayeur et al., *Nat. Neurosci*

4:492-8 (2001); Sainz et al., J. Neurochem 77:896-903 (2001)). Furthermore, dimerization is emerging as a common theme of C-family receptors: the metabotropic glutamate and calcium-sensing receptor are homodimers (Romomano et al., *J. Biol. Chem.* 271:28612-6 (1996); Okamoto et al., *J. Biol. Chem.* 273: 13089-96 (1998); Han et al., *J. Biol. Chem.* 274:100008-13 (1999); Bai et al., *J. Biol. Chem.* 273:23605-10 (1998)), and the structurally related $GABA_B$ receptor is heterodimeric (Jones et al., *Nature* 396:674-9 (1998); Kaupmann et al., *Nature* 396:683-687 (1998); White et al., *Nature* 396: 679-682 (1998); Kuner et al., *Science* 283:74-77 (1999)). The present inventors have demonstrated by functional coexpression of T1Rs in heterologous cells that human T1R2 functions in combination with human T1R3 as a sweet taste receptor and that human T1R1 functions in combination with human T1R3 as an umami taste receptor.

The discoveries discussed herein are especially significant, as previously the development of improved artificial sweeteners has been hampered by the lack of assays for sweet taste. Indeed, the five commonly used commercial artificial sweeteners, all of which activate hT1R2/hT1R3, were discovered serendipitously. Similarly, other than sensory testing, a laborious process, there is no assay for identifying compounds that modulate umami taste. These problems are now alleviated because, as established by experimental results discussed infra, the human sweet and umami receptors have been identified, and assays for these receptors have been developed, particularly assays that use cells that stably express a functional T1R taste receptor, i.e. the sweet or umami taste receptor.

Based thereon the invention provides assays for detecting and characterizing taste–modulating compounds, wherein T1R family members act, as they do in the taste bud, as reporter molecules for the effect on sweet and umami taste of taste-modulating compounds. Particularly provided and within the scope of the invention are assays for identifying compounds that modulate, mimic, enhance and/or block individually, sweet and umami tastes. Methods for assaying the activity of GPCRs, and especially compounds that affect GPCR activity are well known and are applicable to the T1R family member of the present invention and functional combinations thereof. Suitable assays have been identified supra.

The invention also provides compounds that bind T1R1, T1R2, T1R3, T1R2/T1R3 or T1R1/T1R3, or any fragment, portion, or subunit thereof, as disclosed throughout.

In particular, the subject GPCRs can be used in assays to, *e.g.*, measure changes in ligand binding, ion concentration, membrane potential, current flow, ion flux, transcription, receptor-ligand interactions, second messenger concentrations, *in vitro* and *in vivo*. In another embodiment, T1R family members may be recombinantly expressed in cells, and the modulation of taste transduction via GPCR activity may be assayed by measuring changes in Ca^{2+} levels and other intracellular messages such as cAMP, cGMP, or IP_3.

In certain assays, a domain of a T1R polypeptide, *e.g.*, an extracellular, transmembrane, or intracellular domain, is fused to a heterologous polypeptide, thereby forming a chimeric polypeptide, *e.g.*, a chimeric protein with GPCR activity. Particularly contemplated is the use of fragments of T1R1, T1R2 or T1R3 containing the N-terminal ligand-binding domain. Such proteins are useful, *e.g.*, in assays to identify ligands, agonists, antagonists, or other modulators of T1R receptors. For example, a T1R polypeptide can be expressed in a eukaryotic cell as a chimeric receptor with a heterologous, chaperone sequence that facilitates plasma membrane trafficking, or maturation and targeting through the secretory pathway. The optional heterologous sequence may be a PDZ domain-interacting peptide, such as a C-terminal PDZIP fragment (SEQ ID NO 1). PDZIP is an ER export signal, which, according to the present invention, has been shown to facilitate surface expression of heterologous proteins such as the T1R receptors described herein. More particularly, in one aspect of the invention, PDZIP can be used to promote proper targeting of problematic membrane proteins such as olfactory receptors, T2R taste receptors, and the T1R taste receptors described herein.

Examples of such chimeric receptors include trans-species receptors. Any combination of receptor subunits from various species can be used together to form a chimeric receptor, which can then be used to identify tastants, for example. Therefore, contemplated herein is a chimeric T1R2/T1R3 receptor comprising a human T1R2 subunit and a rat T1R3 subunit. Also contemplated is a chimeric T1R2/T1R3 receptor comprising, a rat T1R2 subunit and a human T1R3 subunit. Also contemplated is a chimeric T1R2 receptor subunit comprising, a human extracellular domain, a rat

transmembrane domain and a rat intracellular domain (SEQ ID NOS: 16 and 17, for example). Also contemplated is chimeric T1R3 receptor subunit comprising, a rat extracellular domain, a human transmembrane domain and a human intracellular domain (SEQ ID NOS: 18 and 19, for example.)

Such chimeric T1R receptors can be expressed in any eukaryotic cell, such as HEK-293 cells. Preferably, the cells contain a G protein, preferably a promiscuous G protein such as $G_{\alpha 15}$ or $G_{\alpha 16}$ or another type of promiscuous G protein capable of linking a wide range of GPCRs to an intracellular signaling pathway or to a signaling protein such as phospholipase C. Activation of such chimeric receptors in such cells can be detected using any standard method, such as by detecting changes in intracellular calcium by detecting FURA-2 dependent fluorescence in the cell. If preferred host cells do not express an appropriate G protein, they may be transfected with a gene encoding a promiscuous G protein such as those described in U.S. Application Serial No. 60/243,770, U.S. Application Serial No. 09/984,297, filed October 29, 2001, and U.S. Application Serial No. 09/989,497 filed November 21, 2001 which are herein incorporated by reference in its entirety.

Additional methods of assaying for modulators of taste transduction include *in vitro* ligand-binding assays using: T1R polypeptides, portions thereof, *i.e.*, the extracellular domain, transmembrane region, or combinations thereof, or chimeric proteins comprising one or more domains of a T1R family member; oocyte or tissue culture cells expressing T1R polypeptides, fragments, or fusion proteins; phosphorylation and dephosphorylation of T1R family members; G protein binding to GPCRs; ligand-binding assays; voltage, membrane potential and conductance changes; ion flux assays; changes in intracellular second messengers such as cGMP, cAMP and inositol triphosphate (IP3); and changes in intracellular calcium levels.

Further, the invention provides methods of detecting T1R nucleic acid and protein expression, allowing investigation of taste transduction regulation and specific identification of taste receptor cells. T1R family members also provide useful nucleic acid probes for paternity and forensic investigations. T1R genes are also useful as nucleic acid probes for identifying taste receptor cells, such as foliate, fungiform, circumvallate, geschmackstreifen, and epiglottis taste receptor cells. T1R receptors can

also be used to generate monoclonal and polyclonal antibodies useful for identifying taste receptor cells.

Functionally, the T1R polypeptides comprise a family of related seven transmembrane G protein-coupled receptors, which are believed to be involved in taste transduction and may interact with a G protein to mediate taste signal transduction (*see, e.g.*, Fong, *Cell Signal*, 8:217 (1996); Baldwin, *Curr. Opin. Cell Biol.*, 6:180 (1994)). Structurally, the nucleotide sequences of T1R family members encode related polypeptides comprising an extracellular domain, seven transmembrane domains, and a cytoplasmic domain. Related T1R family genes from other species share at least about 50%, and optionally 60%, 70%, 80%, or 90%, nucleotide sequence identity over a region of at least about 50 nucleotides in length, optionally 100, 200, 500, or more nucleotides in length to the T1R nucleic acid sequences disclosed herein in the Examples, or conservatively modified variants thereof, or encode polypeptides sharing at least about 35 to 50%, and optionally 60%, 70%, 80%, or 90%, amino acid sequence identity over an amino acid region at least about 25 amino acids in length, optionally 50 to 100 amino acids in length to a T1R polypeptide sequence disclosed infra in the Examples conservatively modified variants thereof.

Several consensus amino acid sequences or domains have also been identified that are characteristic of T1R family members. For example, T1R family members typically comprise a sequence having at least about 50%, optionally 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95-99%, or higher, identity to T1R consensus sequences 1 and 2 (SEQ ID NOs. 2 and 3, respectively). These conserved domains thus can be used to identify members of the T1R family, by identity, specific hybridization or amplification, or specific binding by antibodies raised against a domain. T1R consensus sequences include by way of example the following sequences:

T1R Family Consensus Sequence 1: (SEQ ID NO: 2)

(TR)C(FL)(RQP)R(RT)(SPV)(VERKT)FL(AE)(WL)(RHG)E

T1R Family Consensus Sequence 2: (SEQ ID NO: 3)

(LQ)P(EGT)(NRC)YN(RE)A(RK)(CGF)(VLI)T(FL)(AS)(ML)

These consensus sequences are inclusive of those found in the T1R polypeptides described herein, but T1R family members from other organisms may be expected to comprise consensus sequences having about 75% identity or more to the inclusive consensus sequences described specifically herein.

Specific regions of the T1R nucleotide and amino acid sequences may be used to identify polymorphic variants, interspecies homologs, and alleles of T1R family members. This identification can be made *in vitro, e.g.*, under stringent hybridization conditions or PCR (*e.g.*, using primers encoding the T1R consensus sequences identified above), or by using the sequence information in a computer system for comparison with other nucleotide sequences. Different alleles of T1R genes within a single species population will also be useful in determining whether differences in allelic sequences control differences in taste perception between members of the population. Classical PCR-type amplification and cloning techniques are useful for isolating new T1Rs, for example, where degenerate primers are sufficient for detecting related genes across species.

Typically, identification of polymorphic variants and alleles of T1R family members can be made by comparing an amino acid sequence of about 25 amino acids or more, *e.g.*, 50-100 amino acids. Amino acid identity of approximately at least 35 to 50%, and optionally 60%, 70%, 75%, 80%, 85%, 90%, 95-99%, or above typically demonstrates that a protein is a polymorphic variant, interspecies homolog, or allele of a T1R family member. Sequence comparison can be performed using any of the sequence comparison algorithms discussed below. Antibodies that bind specifically to T1R polypeptides or a conserved region thereof can also be used to identify alleles, interspecies homologs, and polymorphic variants.

Polymorphic variants, interspecies homologs, and alleles of T1R genes can be confirmed by examining taste-cell-specific expression of the putative T1R gene or protein. Typically, T1R polypeptides having an amino acid sequence disclosed herein can be used as a positive control in comparison to the putative T1R polypeptide to demonstrate the identification of a polymorphic variant or allele of the T1R family member. The polymorphic variants, alleles, and interspecies homologs are expected to retain the seven transmembrane structure of a G protein-coupled receptor. For further detail, *see* WO 00/06592, which discloses related T1R family members, GPCR-B3s, the contents of which are herein incorporated by reference in a manner consistent with this disclosure. GPCR-B3 receptors are referred to herein as rT1R1 and mT1R1. Additionally, *see* WO 00/06593, which also discloses related T1R family members, GPCR-B4s, the contents of which are herein incorporated by reference in a manner consistent with this disclosure. GPCR-B4 receptors are referred to herein as rT1R2 and

mT1R2. As discussed previously, the invention also includes structure-based assays that utilize the x-ray crystalline structure of a T1R or T1R combination, e.g., hT1R2/hT1R3 or hT1R1/hT1R3, to identify molecules that modulate T1R receptor activity, and thereby modulate sweet and/or umami taste.

The present invention also provides assays, preferably high throughput assays, to identify molecules that enhance, mimic, block and/or modulate T1R receptors. In some assays, a particular domain of a T1R family member is used in combination with a particular domain of another T1R family member, *e.g.*, an extracellular, transmembrane, or intracellular domain or region. In other embodiments, an extracellular domain, transmembrane region or combination thereof may be bound to a solid substrate, and used, *e.g.*, to isolate ligands, agonists, antagonists, or any other molecules that can bind to and/or modulate the activity of a T1R polypeptide.

Various conservative mutations and substitutions are envisioned to be within the scope of the invention. For instance, it is within the level of skill in the art to perform amino acid substitutions using known protocols of recombinant gene technology including PCR, gene cloning, site-directed mutagenesis of cDNA, transfection of host cells, and in-vitro transcription. The variants could then be screened for activity.

Definitions

As used herein, the following terms have the meanings ascribed to them unless specified otherwise.

"Taste cells" include neuroepithelial cells that are organized into groups to form taste buds of the tongue, *e.g.*, foliate, fungiform, and circumvallate cells (*see, e.g.*, Roper et al., Ann. Rev. Neurosci. 12:329-353 (1989)). Taste cells are also found in the palate and other tissues, such as the esophagus and the stomach.

"T1R" refers to one or more members of a family of G protein-coupled receptors that are expressed in taste cells such as foliate, fungiform, and circumvallate cells, as well as cells of the palate, and esophagus (*see, e.g.*, Hoon *et al.*, *Cell*, 96:541-551 (1999), herein incorporated by reference in its entirety). Members of this family are also referred to as GPCR-B3 and TR1 in WO 00/06592 as well as GPCR-B4 and TR2 in WO 00/06593. GPCR-B3 is also herein referred to as rT1R1, and GPCR-B4 is referred to as rT1R2. Taste receptor cells can also be identified on the basis of morphology (*see, e.g.*, Roper, *supra*), or by the expression of proteins specifically

expressed in taste cells. T1R family members may have the ability to act as receptors for sweet taste transduction, or to distinguish between various other taste modalities. Representative T1R sequences, including hT1R1, hT1R2 and hT1R3 are identified infra in the examples.

"T1R" nucleic acids encode a family of GPCRs with seven transmembrane regions that have "G protein-coupled receptor activity," *e.g.*, they may bind to G proteins in response to extracellular stimuli and promote production of second messengers such as IP3, cAMP, cGMP, and Ca^{2+} via stimulation of enzymes such as phospholipase C and adenylate cyclase (for a description of the structure and function of GPCRs, *see, e.g.*, Fong, *supra*, and Baldwin, *supra*). A single taste cell may contain many distinct T1R polypeptides.

The term "T1R" family therefore refers to polymorphic variants, alleles, mutants, and interspecies homologs that: (1) have at least about 35 to 50% amino acid sequence identity, optionally about 60, 75, 80, 85, 90, 95, 96, 97, 98, or 99% amino acid sequence identity to a T1R polypeptide, preferably those identified in Example 1, over a window of about 25 amino acids, optionally 50-100 amino acids; (2) specifically bind to antibodies raised against an immunogen comprising an amino acid sequence preferably selected from the group consisting of the T1R polypeptide sequence disclosed in Example 1 and conservatively modified variants thereof; (3) are encoded by a nucleic acid molecule which specifically hybridize (with a size of at least about 100, optionally at least about 500-1000 nucleotides) under stringent hybridization conditions to a sequence selected from the group consisting of the T1R nucleic acid sequences contained in Example 1, and conservatively modified variants thereof; or (4) comprise a sequence at least about 35 to 50% identical to an amino acid sequence selected from the group consisting of the T1R amino acid sequence identified in Example 1.

Topologically, the TIRs disclosed herein have an "N-terminal domain" also called "extracellular domain" comprising a "venus flytrap domain" and a "cysteine rich domain;" "transmembrane domains" comprising seven transmembrane regions, and corresponding cytoplasmic, and extracellular loops; and a "C-terminal domain" (*see, e.g.*, Hoon *et al., Cell*, 96:541-551 (1999); Buck & Axel, *Cell*, 65:175-187 (1991)). These domains have been structurally identified using methods known to those of skill in the art, such as sequence analysis programs that identify hydrophobic and

hydrophilic domains (Stryer, *Biochemistry*, (3rd ed. 1988). Such domains are useful for making chimeric proteins and for in vitro assays of the invention, *e.g.*, ligand binding assays. The specific binding of a compound to these structurally defined domains provides provides structural definition for the compound.

"Extracellular domains" therefore refers to the domains of T1R polypeptides that protrude from the cellular membrane and are exposed to the extracellular face of the cell. Such domains generally include the "N terminal domain" that is exposed to the extracellular face of the cell, and optionally can include portions of the extracellular loops of the transmembrane domain that are exposed to the extracellular face of the cell, *i.e.*, the loops between transmembrane regions 2 and 3, between transmembrane regions 4 and 5, and between transmembrane regions 6 and 7.

The "N-terminal domain" region starts at the N-terminus and extends to a region close to the start of the first transmembrane domain. More particularly, in one embodiment of the invention, this domain starts at the N-terminus and ends approximately at the conserved glutamic acid at amino acid position 563 plus or minus approximately 20 amino acids. These extracellular domains are useful for *in vitro* ligand-binding assays, both soluble and solid phase. In addition, transmembrane regions, described below, can also bind ligand either in combination with the extracellular domain, and are therefore also useful for *in vitro* ligand-binding assays.

"Cysteine-rich domain" refers to the domain of the polypeptides. This conserved sequence contains several highly-conserved Cys residues that form disulphide bridges, and lies outside the cell membrane.. This region corresponds to the domain of the T1R family members and is found in all three subunits, T1R1-T1R3. The cysteine rich sequence is found in amino acids 510-566 of T1R1, 508-565 of T1R2, and 512-568 or T1R3.

"Transmembrane domain," which comprises the seven "transmembrane regions," refers to the domain of T1R polypeptides that lies within the plasma membrane, and may also include the corresponding cytoplasmic (intracellular) and extracellular loops. In one embodiment, this region corresponds to the domain of T1R family members which starts approximately at the conserved glutamic acid residue at amino acid position 563 plus or minus 20 amino acids and ends approximately at the conserved tyrosine amino acid residue at position 812 plus or minus approximately 10 amino acids. The seven transmembrane regions and extracellular and cytoplasmic loops

can be identified using standard methods, as described in Kyte & Doolittle, *J. Mol. Biol.*, 157:105-32 (1982)), or in Stryer, *supra.*

"Cytoplasmic domains" refers to the domains of T1R polypeptides that face the inside of the cell, *e.g.*, the "C-terminal domain" and the intracellular loops of the transmembrane domain, *e.g.*, the intracellular loop between transmembrane regions 1 and 2, the intracellular loop between transmembrane regions 3 and 4, and the intracellular loop between transmembrane regions 5 and 6.

"C-terminal domain" refers to the region that spans the end of the last transmembrane domain and the C-terminus of the protein, and which is normally located within the cytoplasm. In one embodiment, this region starts at the conserved tyrosine amino acid residue at position 812 plus or minus approximately 10 amino acids and continues to the C-terminus of the polypeptide.

The term "ligand-binding region" or "ligand-binding domain" refers to sequences derived from a taste receptor, particularly a taste receptor that substantially incorporates at least the extracellular domain of the receptor. In one embodiment, the extracellular domain of the ligand-binding region may include the N-terminal domain and, optionally, portions of the transmembrane domain, such as the extracellular loops of the transmembrane domain. The ligand-binding region may be capable of binding a ligand, and more particularly, a compound that enhances, mimics, blocks, and/or modulates taste, e.g., sweet or umami taste.

The phrase "heteromultimer" or "heteromultimeric complex" in the context of the T1R receptors or polypeptides of the invention refers to a functional association of at least one T1R receptor and another receptor, typically another T1R receptor polypeptide (or, alternatively another non-T1R receptor polypeptide). For clarity, the functional co-dependence of the T1Rs is described in this application as reflecting their possible function as heterodimeric taste receptor complexes. However, as discussed previously, functional co-dependence may alternatively reflect an indirect interaction. For example, T1R3 may function solely to facilitate surface expression of T1R1 and T1R2, which may act independently as taste receptors. Alternatively, a functional taste receptor may be comprised solely of T1R3, which is differentially processed under the control of T1R1 or T1R2, analogous to RAMP-dependent processing of the calcium-related receptor.

The phrase "functional effects" in the context of assays for testing compounds that modulate T1R family member mediated taste transduction includes the determination of any parameter that is indirectly or directly under the influence of the receptor, *e.g.*, functional, physical and chemical effects. It includes ligand binding, changes in ion flux, membrane potential, current flow, transcription, G protein binding, GPCR phosphorylation or dephosphorylation, conformation change-based assays, signal transduction, receptor-ligand interactions, second messenger concentrations (*e.g.*, cAMP, cGMP, IP3, or intracellular Ca^{2+}), *in vitro*, *in vivo*, and *ex vivo* and also includes other physiologic effects such increases or decreases of neurotransmitter or hormone release.

By "determining the functional effect" in the context of assays is meant assays for a compound that increases or decreases a parameter that is indirectly or directly under the influence of a T1R family member, *e.g.*, functional, physical and chemical effects. Such functional effects can be measured by any means known to those skilled in the art, *e.g.*, changes in spectroscopic characteristics (*e.g.*, fluorescence, absorbency, refractive index), hydrodynamic (*e.g.*, shape), chromatographic, or solubility properties, patch clamping, voltage-sensitive dyes, whole cell currents, radioisotope efflux, inducible markers, oocyte T1R gene expression; tissue culture cell T1R expression; transcriptional activation of T1R genes; ligand-binding assays; voltage, membrane potential and conductance changes; ion flux assays; changes in intracellular second messengers such as cAMP, cGMP, and inositol triphosphate (IP3); changes in intracellular calcium levels; neurotransmitter release, conformational assays and the like.

A "flavor or tastant" herein refers to a compound or biologically acceptable salt thereof that induces, in a subject, the perception of smell and/or taste, which include sweet, sour, salty, bitter and umami, and others. The subject can be human, animals, and/or a biological assay, such as the ones described and cited in this application.

A "flavor or taste modifier" herein refers to a compound or biologically acceptable salt thereof that modulates, including enhancing or potentiating, inhibiting, and inducing, the smell and/or tastes of a natural or synthetic tastants in a subject.

A "flavor or taste enhancer" herein refers to a compound or biologically acceptable salt thereof that enhances the tastes or smell of a natural or synthetic

tastants, *e.g.*, monosodium glutamate (MSG) for umami taste and fructose for sweet taste.

"Umami tastant" or "umami compound" herein refers to a compound or biologically acceptable salt thereof that elicits a detectable umami taste in a subject, *e.g.*, MSG.

"Sweet tastant" or "sweet compound" herein refers to a compound or biologically acceptable salt thereof that elicits a detectable sweet taste in a subject, *e.g.*, fructose.

An "umami taste modifier" herein refers to a compound or biologically acceptable salt thereof that modulates, including enhancing or potentiating, inhibiting, and inducing, the umami taste of a natural or synthetic umami tastants, *e.g.*, monosodium glutamate (MSG) in a subject.

A "sweet taste modifier" herein refers to a compound or biologically acceptable salt thereof that modulates, including enhancing or potentiating, inhibiting, and inducing, the sweet taste of a natural or synthetic sweet tastants, *e.g.*, fructose, in a subject.

A "taste enhancing amount" herein refers to an amount of a compound that is sufficient to enhance the taste of a natural or synthetic tastants, *e.g.*, monosodium glutamate (MSG) for umami taste or fructose for sweet taste.

"Wet Soup Category" means wet/liquid soups regardless of concentration or container, including frozen Soups. For the purpose of this definition soup(s) means a food prepared from meat, poultry, fish, vegetables, grains, fruit and other ingredients, cooked in a liquid which may include visible pieces of some or all of these ingredients. It may be clear (as a broth) or thick (as a chowder), smooth, pureed or chunky, ready-to-serve, semi-condensed or condensed and may be served hot or cold, as a first course or as the main course of a meal or as a between meal snack (sipped like a beverage). Soup may be used as an ingredient for preparing other meal components and may range from broths (consommé) to sauces (cream or cheese-based soups).

"Dehydrated and Culinary Food Category" means: (i) Cooking aid products such as: powders, granules, pastes, concentrated liquid products, including concentrated bouillon, bouillon and bouillon like products in pressed cubes, tablets or powder or granulated form, which are sold separately as a finished product or as an ingredient within a product, sauces and recipe mixes (regardless of technology); (ii) Meal

solutions products such as: dehydrated and freeze dried soups, including dehydrated soup mixes, dehydrated instant soups, dehydrated ready-to-cook soups, dehydrated or ambient preparations of ready-made dishes, meals and single serve entrées including pasta, potato and rice dishes; and (iii) Meal embellishment products such as: condiments, marinades, salad dressings, salad toppings, dips, breading, batter mixes, shelf stable spreads, barbecue sauces, liquid recipe mixes, concentrates, sauces or sauce mixes, including recipe mixes for salad, sold as a finished product or as an ingredient within a product, whether dehydrated, liquid or frozen.

"Beverage Category" means beverages, beverage mixes and concentrates, including but not limited to, alcoholic and non-alcoholic ready to drink and dry powdered Other examples of foods and beverages wherein compounds according to the invention may be incorporated included by way of example carbonated and non-carbonated beverages, *e.g.*, sodas, juices, alcoholic and non-alcoholic beverages, confectionary products, *e.g.*, cakes, cookies, pies, candies, chewing gums, gelatins, ice creams, sorbets, puddings, jams, jellies, salad dressings, and other condiments, cereal, and other breakfast foods, canned fruits and fruit sauces and the like.

Additionally, the subject compounds can be used in flavor preparations to be added to foods and beverages. In preferred instances the composition will comprise another flavor or taste modifier such as a sweet tastant.

In some instances biologically acceptable salts of the subject compounds may be used. Examples of such salts include alkali and earth metal salts, organic salts, and the like. Specific examples include potassium, sodium, calcium and magnesium salts, hydrochloric or sulfuric acid salts, ethanolamine salts, and the like. The salt will be selected such that it is biologically safe for ingestion and does adversely affect the sweet taste modulatory properties of the compound.

As used herein, the term "medicinal product" includes both solids and liquids which are ingestible non-toxic materials which have medicinal value such as cough syrups, cough drops, aspirin and chewable medicinal tablets. An oral hygiene product includes solids and liquids such as toothpaste or mouthwash.

A "comestibly or medicinally acceptable carrier or excipient" is a medium that is used to prepare a desired dosage form of the inventive compound. A comestibly or medicinally acceptable carrier includes solvents, diluents, or other liquid vehicle;

dispersion or suspension aids; surface active agents; isotonic agents; thickening or emulsifying agents, preservatives; solid binders; lubricants and the like.

"Inhibitors," "activators," "enhancers" and "modulators" of T1R genes or proteins are used to refer to inhibitory, activating, enhancing or modulating molecules identified using *in vitro* and *in vivo* assays for taste transduction, *e.g.*, ligands, agonists, antagonists, and their homologs and mimetics.

Inhibitors are compounds that, *e.g.*, bind to, partially or totally block stimulation, decrease, prevent, delay activation, inactivate, desensitize, or down regulate taste transduction, *e.g.*, antagonists. Activators and enhancers are compounds that, *e.g.*, bind to, enhance, stimulate, increase, open, activate, facilitate, enhance activation, sensitize, or up regulate taste transduction, *e.g.*, agonists. Modulators include compounds that, *e.g.*, alter the interaction of a receptor with: extracellular proteins that bind activators or inhibitor (*e.g.*, ebnerin and other members of the hydrophobic carrier family); G proteins; kinases (*e.g.*, homologs of rhodopsin kinase and beta adrenergic receptor kinases that are involved in deactivation and desensitization of a receptor); and arrestins, which also deactivate and desensitize receptors. Modulators can include genetically modified versions of T1R family members, *e.g.*, with altered activity, as well as naturally occurring and synthetic ligands, antagonists, agonists, small chemical molecules and the like. Such assays for inhibitors and activators include, *e.g.*, expressing T1R family members in cells or cell membranes, applying putative modulator compounds, in the presence or absence of tastants, *e.g.*, sweet tastants, and then determining the functional effects on taste transduction, as described above. Samples or assays comprising T1R family members that are treated with a potential enhancer, activator, inhibitor, or modulator are compared to control samples without the inhibitor, activator, or modulator to examine the extent of modulation. Positive control samples (e.g. a sweet tastant without added modulators) are assigned a relative T1R activity value of 100%.

"EC_{50}" is defined as the amount of a compound that elicits 50% of the maximal response the compound can elicit, whether as an activator, enhancer, or modulator. A dose-dependent response curve was determined for a compound, and the compound concentration corresponding to 50% of the maximal response was derived from the curve, in one example.

"IC_{50}" is defined as the amount of a compound that elicits 50% of the maximal effect the compound can elicit as an inhibitor.

Regarding sweet tastants and enhancers, after a compound is identified, scores of their activities are given as percentage of the maximum fructose intensity (%). In compound dose response, an EC_{50} can be calculated to reflect the potency of the compound as a sweet agonist. In the present invention, an EC_{50} of lower than about 100 mM is indicative of compounds that induce T1R2/T1R3 activity as a sweet agonist. Preferably, a positive hit for a sweet agonist has an EC_{50} value of less than about 1 mM; more preferably less than about a 10 μM.

In sweet enhancement assay experiments, a fructose dose response was run and a second fructose dose response was run with a certain amount of candidate compound at every fructose concentrations at the same time. Then, the EC_{50} ratio can be calculated based on the following definitions:

EC_{50} Ratio = EC_{50} (fructose)/EC_{50} (fructose + [Compound])

wherein "[compound]" refers to the concentration of compound used to elicit (or enhance or potentiate) the fructose dose response. Those concentrations could vary from a pM to an mM, more preferred, from a low nM to μM. A potent sweet enhancer would have a high EC_{50} Ratio at a low concentration of the compound used.

In the present invention, an EC_{50} ratio of greater than 1 is indicative of a compound that modulates (potentiates) T1R2/T1R3 activity and is an sweet enhancer. Preferably, a positive hit will have EC_{50} ratio values of at least 1.20, preferably ranging from at least 1.50 to 100 or even higher.

By contrast, competing agonists (those sweet tastants that bind mutually exclusively) or inhibitors always yield values of EC_{50} ratio less than 1, such as from 0-1.

Regarding umami tastants and enhancers, scores of their activities can be given as percentage of the maximum MSG intensity (%). In compound dose response, an EC_{50} can be calculated to reflect the potency of the compound as umami agonist. In the present invention, an EC_{50} of lower than about 10 mM is indicative of compounds that induce T1R1/T1R3 activity and an umami agonist. Preferably, a positive hit for an umami agonist will have EC_{50} values of less than about 1 mM; more preferably ranging from about a pM to about a low μM.

In enhancement assay experiments, a MSG dose response was run and a second MSG dose response was run with a certain amount of candidate compound at every MSG concentrations at the same time. Then, the EC_{50} ratio is calculated based on the following definitions:

EC_{50} Ratio = EC_{50} (MSG)/EC_{50} (MSG + [Compound])

wherein "[compound]" refers to the concentration of compound used to elicit (or enhance or potentiate) the MSG dose response. Those concentrations can vary from a pM to an mM, more preferred, from a low nM to μM. A potent umami enhancer has a high EC_{50} Ratio at a low concentration of the compound used.

In the present invention, an EC_{50} ratio of greater than 1 is indicative of a compound that modulates (potentiates) T1R1/T1R3 activity and in an umami enhancer. Preferably, a positive hit has EC_{50} ratio values of at least 1.20, preferably ranging from at least 1.50 to 100 or even higher.

Negative control samples (e.g. buffer without an added taste stimulus) are assigned a relative T1R activity value of 0%. Inhibition of a T1R is achieved when a mixture of the positive control sample and a modulator result in the T1R activity value relative to the positive control is about 80%, optionally 50% or 25-0%. Activation of a T1R by a modulator alone is achieved when the T1R activity value relative to the positive control sample is 10%, 25%, 50%, 75%, optionally 100%, optionally 150%, optionally 200-500%, or 1000-3000% higher.

The terms "purified," "substantially purified," and "isolated" as used herein refer to the state of being free of other, dissimilar compounds with which the compound of the invention is normally associated in its natural state, so that the "purified," "substantially purified," and "isolated" subject comprises at least 0.5%, 1%, 5%, 10%, or 20%, and most preferably at least 50% or 75% of the mass, by weight, of a given sample. In one preferred embodiment, these terms refer to the compound of the invention comprising at least 95% of the mass, by weight, of a given sample. As used herein, the terms "purified," "substantially purified," and "isolated," when referring to a nucleic acid or protein, also refers to a state of purification or concentration different than that which occurs naturally in the mammalian, especially human body. Any degree of purification or concentration greater than that which occurs naturally in the mammalian, especially human, body, including (1) the purification from other associated structures or compounds or (2) the association with structures or compounds

to which it is not normally associated in the mammalian, especially human, body, are within the meaning of "isolated." The nucleic acid or protein or classes of nucleic acids or proteins, described herein, may be isolated, or otherwise associated with structures or compounds to which they are not normally associated in nature, according to a variety of methods and processes known to those of skill in the art.

The term "nucleic acid" or "nucleic acid sequence" refers to a deoxy-ribonucleotide or ribonucleotide oligonucleotide in either single- or double-stranded form. The term encompasses nucleic acids, i.e., oligonucleotides, containing known analogs of natural nucleotides. The term also encompasses nucleic-acid-like structures with synthetic backbones (*see e.g., Oligonucleotides and Analogues, a Practical Approach*, ed. F. Eckstein, Oxford Univ. Press (1991); Antisense Strategies, *Annals of the N.Y. Academy of Sciences*, Vol. 600, Eds. Baserga *et al.* (NYAS 1992); Milligan *J. Med. Chem.* 36:1923-1937 (1993); *Antisense Research and Applications* (1993, CRC Press), WO 97/03211; WO 96/39154; Mata, *Toxicol. Appl. Pharmacol.* 144:189-197 (1997); Strauss-Soukup, *Biochemistry* 36:8692-8698 (1997); Samstag, *Antisense Nucleic Acid Drug Dev*, 6:153-156 (1996)).

Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (*e.g.*, degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating, *e.g.*, sequences in which the third position of one or more selected codons is substituted with mixed-base and/or deoxyinosine residues (Batzer *et al.*, *Nucleic Acid Res.*, 19:5081 (1991); Ohtsuka *et al.*, *J. Biol. Chem.*, 260:2605-2608 (1985); Rossolini *et al.*, *Mol. Cell. Probes*, 8:91-98 (1994)). The term nucleic acid is used interchangeably with gene, cDNA, mRNA, oligonucleotide, and polynucleotide.

The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-naturally occurring amino acid polymer.

The term "plasma membrane translocation domain" or simply "translocation domain" means a polypeptide domain that, when incorporated into a polypeptide coding sequence, can with greater efficiency "chaperone" or "translocate" the hybrid

("fusion") protein to the cell plasma membrane than without the domain. For instance, a "translocation domain" may be derived from the amino terminus of the bovine rhodopsin receptor polypeptide, a 7-transmembrane receptor. However, rhodopsin from any mammal may be used, as can other translocation facilitating sequences. Thus, the translocation domain is particularly efficient in translocating 7-transmembrane fusion proteins to the plasma membrane, and a protein (*e.g.*, a taste receptor polypeptide) comprising an amino terminal translocating domain will be transported to the plasma membrane more efficiently than without the domain. However, if the N-terminal domain of the polypeptide is active in binding, as with the T1R receptors of the present invention, the use of other translocation domains may be preferred. For instance, a PDZ domain-interacting peptide, as described herein, may be used.

The "translocation domain," "ligand-binding domain", and chimeric receptors compositions described herein also include "analogs," or "conservative variants" and "mimetics" ("peptidomimetics") with structures and activity that substantially correspond to the exemplary sequences. Thus, the terms "conservative variant" or "analog" or "mimetic" refer to a polypeptide which has a modified amino acid sequence, such that the change(s) do not substantially alter the polypeptide's (the conservative variant's) structure and/or activity, as defined herein. These include conservatively modified variations of an amino acid sequence, *i.e.*, amino acid substitutions, additions or deletions of those residues that are not critical for protein activity, or substitution of amino acids with residues having similar properties (*e.g.*, acidic, basic, positively or negatively charged, polar or non-polar, *etc.*) such that the substitutions of even critical amino acids does not substantially alter structure and/or activity.

More particularly, "conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein.

For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can

be altered to any of the corresponding codons described without altering the encoded polypeptide.

Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein, which encodes a polypeptide, also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid, which encodes a polypeptide, is implicit in each described sequence.

Conservative substitution tables providing functionally similar amino acids are well known in the art. For example, one exemplary guideline to select conservative substitutions includes (original residue followed by exemplary substitution): ala/gly or ser; arg/lys; asn/gln or his; asp/glu; cys/ser; gln/asn; gly/asp; gly/ala or pro; his/asn or gln; ile/leu or val; leu/ile or val; lys/arg or gln or glu; met/leu or tyr or ile; phe/met or leu or tyr; ser/thr; thr/ser; trp/tyr; tyr/trp or phe; val/ile or leu. An alternative exemplary guideline uses the following six groups, each containing amino acids that are conservative substitutions for one another: 1) Alanine (A), Serine (S), Threonine (T); 2) Aspartic acid (D), Glutamic acid (E); 3) Asparagine (N), Glutamine (Q); 4) Arginine (R), Lysine (I); 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W); (*see also, e.g.*, Creighton, *Proteins*, W.H. Freeman and Company (1984); Schultz and Schimer, *Principles of Protein Structure*, Springer-Vrlag (1979)). One of skill in the art will appreciate that the above-identified substitutions are not the only possible conservative substitutions. For example, for some purposes, one may regard all charged amino acids as conservative substitutions for each other whether they are positive or negative. In addition, individual substitutions, deletions or additions that alter, add or delete a single amino acid or a small percentage of amino acids in an encoded sequence can also be considered "conservatively modified variations."

The terms "mimetic" and "peptidomimetic" refer to a synthetic chemical compound that has substantially the same structural and/or functional characteristics of the polypeptides, *e.g.*, translocation domains, ligand-binding domains, or chimeric receptors of the invention. The mimetic can be either entirely composed of synthetic,

non-natural analogs of amino acids, or may be a chimeric molecule of partly natural peptide amino acids and partly non-natural analogs of amino acids. The mimetic can also incorporate any amount of natural amino acid conservative substitutions as long as such substitutions also do not substantially alter the mimetic's structure and/or activity.

As with polypeptides of the invention which are conservative variants, routine experimentation will determine whether a mimetic is within the scope of the invention, *i.e.*, that its structure and/or function is not substantially altered. Polypeptide mimetic compositions can contain any combination of non-natural structural components, which are typically from three structural groups: a) residue linkage groups other than the natural amide bond ("peptide bond") linkages; b) non-natural residues in place of naturally occurring amino acid residues; or c) residues which induce secondary structural mimicry, i.e., to induce or stabilize a secondary structure, *e.g.*, a beta turn, gamma turn, beta sheet, alpha helix conformation, and the like. A polypeptide can be characterized as a mimetic when all or some of its residues are joined by chemical means other than natural peptide bonds. Individual peptidomimetic residues can be joined by peptide bonds, other chemical bonds or coupling means, such as, *e.g.*, glutaraldehyde, N-hydroxysuccinimide esters, bifunctional maleimides, N,N'-dicyclohexylcarbodiimide (DCC) or N,N'-diisopropylcarbodiimide (DIC). Linking groups that can be an alternative to the traditional amide bond ("peptide bond") linkages include, *e.g.*, ketomethylene (*e.g.*, -C(=O)-CH_2- for -C(=O)-NH-), aminomethylene (CH_2-NH), ethylene, olefin (CH=CH), ether (CH_2-O), thioether (CH_2-S), tetrazole (CN_4), thiazole, retroamide, or ester (*see*, *e.g.*, Spatola, *Chemistry and Biochemistry of Amino Acids, Peptides and Proteins*, Vol. 7, pp 267-357, "Peptide Backbone Modifications," Marcell Dekker, NY (1983)). A polypeptide can also be characterized as a mimetic by containing all or some non-natural residues in place of naturally occurring amino acid residues; non-natural residues are well described in the scientific and patent literature.

A "label" or a "detectable moiety" is a composition detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include ^{32}P, fluorescent dyes, electron-dense reagents, enzymes (*e.g.*, as commonly used in an ELISA), biotin, digoxigenin, or haptens and proteins which can be made detectable, *e.g.*, by incorporating a radiolabel into the peptide or used to detect antibodies specifically reactive with the peptide.

A "labeled nucleic acid probe or oligonucleotide" is one that is bound, either covalently, through a linker or a chemical bond, or noncovalently, through ionic, van der Waals, electrostatic, or hydrogen bonds to a label such that the presence of the probe may be detected by detecting the presence of the label bound to the probe.

As used herein a "nucleic acid probe or oligonucleotide" is defined as a nucleic acid capable of binding to a target nucleic acid of complementary sequence through one or more types of chemical bonds, usually through complementary base pairing, usually through hydrogen bond formation. As used herein, a probe may include natural (i.e., A, G, C, or T) or modified bases (7-deazaguanosine, inosine, *etc.*). In addition, the bases in a probe may be joined by a linkage other than a phosphodiester bond, so long as it does not interfere with hybridization. Thus, for example, probes may be peptide nucleic acids in which the constituent bases are joined by peptide bonds rather than phosphodiester linkages. It will be understood by one of skill in the art that probes may bind target sequences lacking complete complementarity with the probe sequence depending upon the stringency of the hybridization conditions. The probes are optionally directly labeled as with isotopes, chromophores, lumiphores, chromogens, or indirectly labeled such as with biotin to which a streptavidin complex may later bind. By assaying for the presence or absence of the probe, one can detect the presence or absence of the select sequence or subsequence.

The term "heterologous" when used with reference to portions of a nucleic acid indicates that the nucleic acid comprises two or more subsequences that are not found in the same relationship to each other in nature. For instance, the nucleic acid is typically recombinantly produced, having two or more sequences from unrelated genes arranged to make a new functional nucleic acid, *e.g.*, a promoter from one source and a coding region from another source. Similarly, a heterologous protein indicates that the protein comprises two or more subsequences that are not found in the same relationship to each other in nature (*e.g.*, a fusion protein).

A "promoter" is defined as an array of nucleic acid sequences that direct transcription of a nucleic acid. As used herein, a promoter includes necessary nucleic acid sequences near the start site of transcription, such as, in the case of a polymerase II type promoter, a TATA element. A promoter also optionally includes distal enhancer or repressor elements, which can be located as much as several thousand base pairs

from the start site of transcription. A "constitutive"promoter is a promoter that is active under most environmental and developmental conditions.

An "inducible" promoter is a promoter that is active under environmental or developmental regulation. The term "operably linked" refers to a functional linkage between a nucleic acid expression control sequence (such as a promoter, or array of transcription factor binding sites) and a second nucleic acid sequence, wherein the expression control sequence directs transcription of the nucleic acid corresponding to the second sequence.

As used herein, "recombinant" refers to a polynucleotide synthesized or otherwise manipulated *in vitro* (*e.g.*, "recombinant polynucleotide"), to methods of using recombinant polynucleotides to produce gene products in cells or other biological systems, or to a polypeptide ("recombinant protein") encoded by a recombinant polynucleotide. "Recombinant means" also encompass the ligation of nucleic acids having various coding regions or domains or promoter sequences from different sources into an expression cassette or vector for expression of, *e.g.*, inducible or constitutive expression of a fusion protein comprising a translocation domain of the invention and a nucleic acid sequence amplified using a primer of the invention.

As used herein, a "stable cell line" refers to a cell line, which stably, i.e. over a prolonged period, expresses a heterologous nucleic sequence, i.e. a T1R or G protein. In preferred embodiments, such stable cell lines will be produced by transfecting appropriate cells, typically mammalian cells, e.g. HEK-293 cells, with a linearized vector that contains a T1R expression construct, i.e. T1R1, T1R2 and/or T1R3. Most preferably, such stable cell lines will be produced by co-transfecting two linearized plasmids that express hT1R1 and hT1R3 or hT1R2 and hT1R3 and an appropriate selection procedure to generate cell lines having these genes stably integrated therein. Most preferably, the cell line will also stably express a G protein such as $G_{\alpha 15}$.

The phrase "selectively (or specifically) hybridizes to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent hybridization conditions when that sequence is present in a complex mixture (*e.g.*, total cellular or library DNA or RNA).

The phrase "stringent hybridization conditions" refers to conditions under which a probe will hybridize to its target subsequence, typically in a complex mixture of nucleic acid, but to no other sequences. Stringent conditions are sequence dependent

and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Techniques in Biochemistry and Molecular Biology - Hybridization with Nucleic Probes*, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally, stringent conditions are selected to be about 5-10° C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength pH. The Tm is the temperature (under defined ionic strength, pH, and nucleic concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at Tm, 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30° C for short probes (*e.g.*, 10 to 50 nucleotides) and at least about 60° C for long probes (*e.g.*, greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. For selective or specific hybridization, a positive signal is at least two times background, optionally 10 times background hybridization. Exemplary stringent hybridization conditions can be as following: 50% formamide, 5x SSC, and 1% SDS, incubating at 42°C, or, 5x SSC, 1% SDS, incubating at 65°C, with wash in 0.2x SSC, and 0.1 % SDS at 65°C. Such hybridizations and wash steps can be carried out for, *e.g.*, 1, 2, 5, 10, 15, 30, 60; or more minutes.

Nucleic acids that do not hybridize to each other under stringent conditions are still substantially related if the polypeptides that they encode are substantially related. This occurs, for example, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. In such cases, the nucleic acids typically hybridize under moderately stringent hybridization conditions. Exemplary "moderately stringent hybridization conditions" include a hybridization in a buffer of 40% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 1X SSC at 45°C. Such hybridizations and wash steps can be carried out for, *e.g.*, 1, 2, 5, 10, 15, 30, 60, or more minutes. A positive hybridization is at least twice background. Those of ordinary skill will readily recognize that alternative hybridization and wash conditions can be utilized to provide conditions of similar stringency.

"Antibody" refers to a polypeptide comprising a framework region from an immunoglobulin gene or fragments thereof that specifically binds and recognizes an antigen. The recognized immunoglobulin genes include the kappa, lambda, alpha, gamma, delta, epsilon, and mu constant region genes, as well as the myriad immunoglobulin variable region genes. Light chains are classified as either kappa or lambda. Heavy chains are classified as gamma, mu, alpha, delta, or epsilon, which in turn define the immunoglobulin classes, IgG, IgM, IgA, IgD and IgE, respectively.

An exemplary immunoglobulin (antibody) structural unit comprises a tetramer. Each tetramer is composed of two identical pairs of polypeptide chains, each pair having one "light" (about 25 kDa) and one "heavy" chain (about 50-70 kDa). The N-terminus of each chain defines a variable region of about 100 to 110 or more amino acids primarily responsible for antigen recognition. The terms "variable light chain" (VL) and "variable heavy chain" (VH) refer to these light and heavy chains respectively.

A "chimeric antibody" is an antibody molecule in which (a) the constant region, or a portion thereof, is altered, replaced or exchanged so that the antigen binding site (variable region) is linked to a constant region of a different or altered class, effector function and/or species, or an entirely different molecule which confers new properties to the chimeric antibody, *e.g.*, an enzyme, toxin, hormone, growth factor, drug, *etc.*; or (b) the variable region, or a portion thereof, is altered, replaced or exchanged with a variable region having a different or altered antigen specificity.

An "anti-T1R" antibody is an antibody or antibody fragment that specifically binds a polypeptide encoded by a T1R gene, cDNA, or a subsequence thereof.

The term "immunoassay" is an assay that uses an antibody to specifically bind an antigen. The immunoassay is characterized by the use of specific binding properties of a particular antibody to isolate, target, and/or quantify the antigen.

The phrase "specifically (or selectively) binds" or "specifically (or selectively) reacts with," when referring to a molecule or composition, refers to a binding reaction that is determinative of the presence of the molecule in a heterogeneous population of other biologics. Thus, under designated conditions, the specified molecules bind to a particular receptor at least two times the background and do not substantially bind in a significant amount to other molecules present in the sample. Specific binding to a

receptor under such conditions may require a receptor that is selected for its specificity for a particular molecule.

Regarding antibodies, a variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays are routinely used to select antibodies specifically immunoreactive with a protein (*see, e.g.*, Harlow & Lane, *Antibodies, A Laboratory Manual*, (1988), for a description of immunoassay formats and conditions that can be used to determine specific immunoreactivity). Typically a specific or selective reaction will be at least twice background signal or noise and more typically more than 10 to 100 times background.

The phrase "selectively associates with" refers to the ability of a nucleic acid to "selectively hybridize" with another as defined above, or the ability of an antibody to "selectively (or specifically) bind to a protein, as defined above.

The term "expression vector" refers to any recombinant expression system for the purpose of expressing a nucleic acid sequence of the invention *in vitro* or *in vivo*, constitutively or inducibly, in any cell, including prokaryotic, yeast, fungal, plant, insect or mammalian cell. The term includes linear or circular expression systems. The term includes expression systems that remain episomal or integrate into the host cell genome. The expression systems can have the ability to self-replicate or not, i.e., drive only transient expression in a cell. The term includes recombinant expression "cassettes which contain only the minimum elements needed for transcription of the recombinant nucleic acid.

By "host cell" is meant a cell that contains an expression vector and supports the replication or expression of the expression vector. Host cells may be prokaryotic cells such as E. coli, or eukaryotic cells such as yeast, insect, amphibian, worm or mammalian cells such as CHO, Hela, HEK-293, and the like, *e.g.*, cultured cells, explants, and cells *in vivo*.

Compounds

As discussed above, there are different domains on the T1R receptors. T1R1, T1R2, and T1R3 each contain an N-terminal extracellular domain (also known as the Venus flytrap domain), transmembrane domains comprising seven transmembrane regions, and corresponding cytoplasmic, and extracellular loops; a cysteine rich

domain, and a C-terminal domain. Each region defines a specific set of compounds that bind specifically to that region.

In humans, the N-terminal extracellular domain comprises amino acids 1 to 560 of hT1R2 and amino acids 1 to 563 of hT1R3. In rats, the N-terminal extracellular domain comprises amino acids 1 to 564 of rT1R2, and amino acids 1 to 568 of rT1R3.

In humans, the C-terminal transmembrane domain and intracellular domain comprise amino acids 561 to 839 of hT1R2, and amino acids 564 to 852 of hT1R3. In rats, the C-terminal transmembrane domain and intracellular domain comprise amino acids 565 to 842 of rT1R2, and amino acids 569 to 858 of rT1R3.

Metabotropic glutamate receptors (mGluR) are another class of C-class G protein-coupled receptors that respond to glutamate. These are found primarily in the brain and neuronal tissue where they play a role in neuronal signaling. The mGluR N-terminal extracellular domain can be covalently linked to a T1R in order to create chimeric receptors. The mGluR receptor can be any of mGluR1-mGluR8, for example. Different ligands bind to different domains on different subunits of both the umami and the sweet receptors. For example, aspartame and neotame bind to the N-terminal extracellular domain of T1R2, while cyclamate, neohesperidin dihydrochalcone (NHDC), and lactisole bind to the transmembrane domain of T1R3. Because T1R3 is one of the two subunits in the T1R1/T1R3 umami taste receptor, cyclamate, NHDC and lactisole can interact with T1R3 in the T1R1/T1R3 umami taste receptor as well. Cyclamate and NHDC enhance the activity of the umami taste receptor, while lactisole inhibits the umami receptor.

The specific binding compounds of the invention as it relates to umami tastants comprise amides. The amide compounds also comprise certain sub-classes of amide derivatives or classes of derivatives related to amides, such as for example ureas, urethanes, oxalamides, acrylamides, and the like.

Molecules that interact with the transmembrane domain of T1R2, for example, can be modulators of sweet taste, and molecules that interact with the transmembrane domain of T1R3 can be modulators of sweet taste and/or umami taste.

Human T1R2/T1R3 recognizes a group of sweeteners which are not recognized by rat T1R2/T1R3, and human but not rat T1R2/T1R3 is inhibited by lactisole. When the extracellular domain of human T1R2 was replaced by its rat counterpart, the human

receptor lost the ability to recognize aspartame, indicating that this part of human T1R2 is required for binding to aspartame. Inversely, when the extracellular domain of rat T1R2 was replaced by its human counterpart, the rat receptor acquired the ability to recognize aspartame, indicating that this part of the human T1R2 is sufficient to bind aspartame. By the same principle, the transmembrane domain of human T1R3 was required and sufficient for

Table 6 shows the abbreviations used to represent various rat/human chimeric receptors and receptor subunits.

TABLE 6

hT1R2 – human T1R2
hT1R3 – human T1R3
rT1R2 – rat T1R2
rT1R3 – rat T1R3
hT1R2/rT1R3 – a receptor composed of human T1R2 and rat T1R3
rT1R2/hT1R3 - a receptor composed of a rat T1R2 and human T1R3
hT1R2/h3-r3 – a receptor composed of human T1R2 and a chimeric T1R3 with human N-terminal extracellular domain and rat transmembrane and C-terminal domain
rT1R2/r3-h3 – a receptor composed of rat T1R2 and a chimeric T1R3 with rat N-terminal extracellular domain and human transmembrane and C-terminal domain
h2-r2/rT1R3 – a receptor composed of a chimeric T1R2 with human N-terminal extracellular domain and rat transmembrane and C-terminal domain and rat T1R3
r2-h2/rT1R3 – a receptor composed of a chimeric T1R2 with rat N-terminal extracellular domain and human transmembrane and C-terminal domain and rat T1R3
h2-h1/hT1R3 – a receptor composed of a chimeric T1R with human T1R2 N-terminal extracellular domain and human T1R1 transmembrane and C-terminal domain and human T1R3
h1-h2/hT1R3 – a receptor composed of a chimeric T1R with human T1R1 N-terminal extracellular domain and human T1R2 transmembrane and C-terminal domain and human T1R3
h2-mGluR1/h3-mGluR1 – a receptor composed of a N-terminal extracellular domain from hT1R2 covalently linked to the transmembrane and C-terminal domain of mGluR1 and a N-terminal extracellular domain from hT1R3 covalently linked to the transmembrane and C-terminal domain of mGluR1 h1-mGlu1R/h3-mGluR1 – a receptor composed of a N-terminal extracellular domain from hT1R1 covalently linked to the transmembrane and C-terminal domain of mGluR1 and a N-terminal extracellular domain from hT1R3 covalently linked to the transmembrane and

C-terminal domain of mGluR1
mGluR1-h2/mGluR1-h3 – a receptor composed of a N-terminal extracellular domain from mGluR1 covalently linked to the transmembrane and C-terminal domain of hT1R2 and a N-terminal extracellular domain from a mGluR1 covalently linked to the transmembrane and C-terminal domain of hT1R3
mGluR1-h1/mGluR1-h3 – receptor composed of a N-terminal extracellular domain from mGluR1 covalently linked to the transmembrane and C-terminal domain of hT1R1 and a N-terminal extracellular domain from mGluR1 covalently linked to the transmembrane and C-terminal domain of hT1R3

Disclosed herein are non-naturally occurring compounds that specifically bind to the T1R2/T1R3 receptor comprising hT1R2/hT1R3 but not rT1R2/rT1R3. Examples of such compounds include, but are not limited to neotame, aspartame, cyclamate, lactisol, Compound 883360, Compound 6542888, Compound 403249, Compound 6364395, Dihydroxybenzoic acid (DHB), Compound 6542888, and neohesperidine dihydrochalcone (NHDC) Additional examples are found in Tables 1-4. The organic, non-peptide compounds can be approximately the size of a box of dimensions 15x8x8 angstroms, more preferably the dimension should be 12x5x5 angstroms.

Also disclosed are compounds that specifically bind to a T1R2/T1R3 receptor comprising hT1R2/rT1R3 but not rT1R2/hT1R3. Examples of such compounds include, but are not limited to aspartame, and neotame. Additional examples are found in Table 5.

Also disclosed are compounds that specifically bind to the N-terminal extracellular domain of T1R2 of the hT1R2/hT1R3 receptor. Examples of such compounds include, but are not limited to neotame, aspartame carbohydrate sugars (e.g. sucrose, fructose, glucose, tagatose, erythritol, sorbitol, maltose, xylitol, lactose and galactose, as well as all other carbohydrate sugars). Additional examples are found in Table 5.

Also disclosed are compounds that specifically bind to the Venus Flytrap Domain (VFD) of T1R2 of the hT1R2/hT1R3 and hT1R2/rT1R3 receptor.

Also disclosed are compounds that specifically bind to the N-terminal Venus flytrap domain of the T1R2 subunit of the T1R2/T1R3 receptor. More specifically, also disclosed are compounds that specifically bind to amino acid residues 144 and 302 of the human N-terminal Venus flytrap domain of the T1R2 subunit of theT1R2/T1R3

receptor. Examples of such compounds include, but are not limited to aspartame, neotame, carbohydrates, and sweet amino acids, such as D-Trp, Ala, and Gly.

Also disclosed are compounds that specifically bind to the cysteine-rich region of T1R2 of the hT1R2/hT1R3 receptor. Also disclosed are compounds that specifically bind to the Transmembrane Domain (TM) of T1R2 of the hT1R2/hT1R3 receptor.

Also disclosed are compounds that specifically bind to a T1R2/T1R3 receptor comprising rT1R2/hT1R3 but not hT1R2/rT1R3. Examples of such compounds include, but are not limited to cyclamate, NHDC, lactisole, Compound 883360, Compound 403249, and Compound 6364395. Additional examples are found in Table 5.

Also disclosed are compounds that specifically bind to hT1R2/hT1R3 and rT1R2/r3-h3 but not to rT1R2/rT1R3 or to hT1R2/h3-r3. Examples of such compounds include, but are not limited to cyclamate, NHDC, lactisole, Compound 883360, Compound 403249 and Compound 6364395.

Also disclosed are compounds that specifically bind to extracellular loop 2 and extracellular loop 3 of the human C-terminal domain of the T1R3 subunit of theT1R2/T1R3 receptor. Also disclosed are compounds that specifically bind to hT1R2/hT1R3 and r2-h2/rT1R3 but not to rT1R2/rT1R3 or to h2-r2/hT1R3.

Also disclosed are compounds that specifically bind to the human N-terminal extracellular domain of the T1R3 subunit of the T1R2/T1R3 receptor. Also disclosed are compounds that specifically bind to the Venus Flytrap Domain (VFD) of T1R3 of the hT1R2/hT1R3 receptor. Examples of such compounds include, but are not limited to aspartame, neotame, carbohydrates, and sweet amino acids, such as D-Trp, Ala, and Gly.

Also disclosed are compounds that specifically bind to the Transmembrane Domain of T1R3 of the hT1R2/hT1R3 receptor. Also disclosed are compounds that specifically bind to extracellular loop 2 and extracellular loop 3 of the human transmembrane domain of the T1R3 subunit of T1R2/T1R3. Examples of such compounds include, but are not limited to cyclamate.

The compound of the invention does not include sucrose, fructose, glucose, erythritol, isomalt, lactitol, mannitol, sorbitol, xylitol, certain known natural terpenoids, flavonoids, or protein sweeteners, di-peptides, tri-peptides, aspartame, saccharin, sucralose, halogenated saccharides, acesulfame-K, cyclamate, sucralose, and alitame.

neotame, perillartine, SC-45647, SC-40014, monellin, NC-002740-01, thaumatin, CC-00100, NC-00420, alitame, SC-44102, dulcin, NC-00576, slycyrrhizic Acid, stevioside, Na-Saccharin, D-tryptophan, cyclamate, DHB, glycolic Acid, glycine, D (-)fructose, homofuronol, D (-) tagatose, maltose, D (+) glucose, D-sorbitol, D (+) galactose, α-lactose, L()fructose, L (+) Compound 403249, and glucose.

Optionally, a compound of the invention is also not Compound 6364395.

Also disclosed herein are compounds that bind a truncated region of a T1R domain. For example, disclosed are compounds that specifically bind to the TM domain of T1R2 of a truncated sweet receptor comprising h2TM/h3TM, compounds that specifically bind to the TM domain of T1R3 of a truncated sweet receptor comprising h2TM/h3TM, compounds that specifically bind to the TM domain of T1R2 of a chimeric receptor comprising mGluR-h2/mGluR-h3, compounds that specifically bind to the TM domain of T1R3 of a chimeric receptor comprising mGluR-h2/mGluR-h3, compounds that binds to the TM domain of T1R1 of a truncated savory receptor comprising h1TM/h3TM, compound that binds to the TM domain of T1R3 of a truncated sweet receptor comprising h1TM/h3TM, compounds that bind to the TM domain of T1R1 of a chimeric receptor comprising mGluR-h1/mGluR-h3, and compounds that bind to the TM domain of T1R3 of a chimeric receptor comprising mGluR-h1/mGluR-h3. SEQ ID NOS: 29-33 represent these truncated receptors.

The compounds of the invention do not include monosodium glutamate ("MSG"), inosine monophosphate (IMP) or guanosine monophosphate (GMP), sucrose, fructose, glucose, erythritol, isomalt, lactitol, mannitol, sorbitol, xylitol, certain known natural terpenoids, flavonoids, or protein sweeteners, di-peptides, tri-peptides aspartame, saccharin, sucralose, halogenated saccharides, acesulfame-K, cyclamate, sucralose, alitame, monosodium glutamate ("MSG"), inosine monophosphate (IMP) or guanosine monophosphate (GMP), or adenosine monophospate.

Compound 403249 is (5-(4H-benzo[d][1,3]oxathiin-2-yl)-2-methyoxyphenol, while Compound 6364395 is 3-(3-hydroxy-4-methoxyphenethyl)benzo[d]isoxyazole-4,6-diol.

The compounds described above can demonstrate a compound-dependent increase in fluorescence with an activity compared to the maximal activity for fructose of at least 25% in a fluorescence-based assay using a FLIPR instrument (Fluorometric Intensity Plate Reader, Molecular Devices, Sunnyvale, CA). For examples of this

protocol, see Examples 12 and 18. The compounds can also demonstrate a compound-dependent decrease in the EC50 for a sweetener by at least two-fold in a fluorescence-based assay using a FLIPR (Molecular Devices) instrument. Furthermore, in a cell-based assay, the compound can result in at least 10 out of 100 cells transfected with wild-type or chimeric receptor showing a compound-dependent increase in fluorescence. An example of a cell-based assay can be found in Example 24. The compound can also demonstrate a compound-dependent increase of at least 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, two-fold or greater, or any number in between, in the number of fluorescent cells in response to a sub-maximal level of a sweetener. The response can be measured by fluorescence, calcium levels, IP3 levels, cAMP levels, GTPγS binding, or reporter gene activity (e.g. luciferase, beta-galactosidase).

Furthermore, the compounds disclosed herein can have one or more of the following characteristics in a cell: a decreased EC_{50} compared to a control of at least approximately 50%, increased intracellular Ca^{2+} level by at least approximately 25%, increased intracellular cAMP by at least approximately 25%, increased intracellular cGMP by at least approximately 25%, increased intracellular IP_3 by at least approximately 25%, or increased G protein binding of GTPγS by at least approximately 25%.

Methods of Using the Compounds

Also disclosed are methods modulating the savory taste of a comestible or medicinal product comprising providing at least one comestible or medicinal product, or a precursor thereof, and combining the comestible or medicinal product or precursor thereof with at least a savory flavor modulating amount of at least one non-naturally occurring compound as disclosed herein, or a comestibly acceptable salt thereof, so as to form a modified comestible or medicinal product; thereby modulating the savory taste of a comestible or medicinal product.

Also disclosed are methods for inhibiting the savory taste of a comestible or medicinal product comprising providing at least one comestible or medicinal product, or a precursor thereof, and combining the comestible or medicinal product or precursor thereof with at least a savory flavor inhibiting amount of at least one non-naturally occurring compound as disclosed herein, or a comestibly acceptable salt thereof, so as

to form a modified comestible or medicinal product; thereby inhibiting the savory taste of a comestible or medicinal product.

Also disclosed are methods for increasing the savory taste of a comestible or medicinal product comprising providing at least one comestible or medicinal product, or a precursor thereof, and combining the comestible or medicinal product or precursor thereof with at least a savory flavor increasing amount of at least one non-naturally occurring compound as disclosed herein, or a comestibly acceptable salt thereof, so as to form a modified comestible or medicinal product; thereby increasing the savory taste of a comestible or medicinal product.

Also disclosed are methods for modulating the sweet taste of a comestible or medicinal product comprising providing at least one comestible or medicinal product, or a precursor thereof, and combining the comestible or medicinal product or precursor thereof with at least a sweet flavor modulating amount of at least one non-naturally occurring compound as disclosed herein, or a comestibly acceptable salt thereof, so as to form a modified comestible or medicinal product; thereby modulating the sweet taste of a comestible or medicinal product.

Also disclosed are methods for inhibiting the sweet taste of a comestible or medicinal product comprising providing at least one comestible or medicinal product, or a precursor thereof, and combining the comestible or medicinal product or precursor thereof with at least a sweet flavor inhibiting amount of at least one non-naturally occurring compound as disclosed herein, or a comestibly acceptable salt thereof, so as to form a modified comestible or medicinal product; thereby inhibiting the sweet taste of a comestible or medicinal product.

Also disclosed are methods for increasing the sweet taste of a comestible or medicinal product comprising providing at least one comestible or medicinal product, or a precursor thereof, and combining the comestible or medicinal product or precursor thereof with at least a sweet flavor increasing amount of at least one non-naturally occurring compound as disclosed herein, or a comestibly acceptable salt thereof, so as to form a modified comestible or medicinal product; thereby increasing the sweet taste of a comestible or medicinal product.

Also disclosed are methods of enhancing umami taste perception comprising contacting an umami receptor with cyclamate and NHDC, and their derivatives, as well

as methods of enhancing umami taste perception comprising contacting an umami receptor with lactisole derivatives. Also disclosed are methods of enhancing sweet taste perception comprising contacting an sweet receptor with cyclamate and NHDC, and their derivatives. Also disclosed are methods of enhancing sweet taste perception comprising contacting an sweet receptor with lactisole derivatives.

Isolation and Expression of T1R Polypeptides

Isolation and expression of the T1Rs, or fragments or variants thereof, of the invention can be performed as described below. PCR primers can be used for the amplification of nucleic acids encoding taste receptor ligand-binding regions, and libraries of these nucleic acids can optionally be generated. Individual expression vectors or libraries of expression vectors can then be used to infect or transfect host cells for the functional expression of these nucleic acids or libraries. These genes and vectors can be made and expressed *in vitro* or *in vivo*. One of skill will recognize that desired phenotypes for altering and controlling nucleic acid expression can be obtained by modulating the expression or activity of the genes and nucleic acids (*e.g.*, promoters, enhancers and the like) within the vectors of the invention. Any of the known methods described for increasing or decreasing expression or activity can be used. The invention can be practiced in conjunction with any method or protocol known in the art, which are well described in the scientific and patent literature.

The nucleic acid sequences of the invention and other nucleic acids used to practice this invention, whether RNA, cDNA, genomic DNA, vectors, viruses or hybrids thereof, may be isolated from a variety of sources, genetically engineered, amplified, and/or expressed recombinantly. Any recombinant expression system can be used, including, in addition to mammalian cells, *e.g.*, bacterial, yeast, insect, or plant systems.

Alternatively, these nucleic acids can be synthesized *in vitro* by well-known chemical synthesis techniques, as described in, *e.g.*, Carruthers, *Cold Spring Harbor Symp. Quant. Biol.* 47:411-418 (1982); Adams, *Am. Chem. Soc.* 105:661 (1983); Belousov, *Nucleic Acids Res.* 25:3440-3444 (1997); Frenkel, *Free Radic. Biol. Med.* 19:373-380 (1995); Blommers, *Biochemistry* 33:7886-7896 (1994); Narang, *Meth. Enzymol.* 68:90 (1979); Brown, *Meth. Enzymol.* 68:109 (1979); Beaucage, *Tetra. Lett.* 22:1859 (1981); U.S. Patent No. 4,458,066. Double-stranded DNA fragments may

then be obtained either by synthesizing the complementary strand and annealing the strands together under appropriate conditions, or by adding the complementary strand using DNA polymerase with an appropriate primer sequence.

Techniques for the manipulation of nucleic acids, such as, for example, for generating mutations in sequences, subcloning, labeling probes, sequencing, hybridization and the like are well described in the scientific and patent literature. *See, e.g.*, Sambrook, ed., Molecular Cloning: a Laboratory manual (2nd ed.), Vols. 1-3, Cold Spring Harbor Laboratory (1989); Current Protocols in Molecular Biology, Ausubel, ed. John Wiley & Sons, Inc., New York (1997); Laboratory Techniques in Biochemistry and Molecular Biology: Hybridization With Nucleic Acid Probes, Part I, Theory and Nucleic Acid Preparation, Tijssen, ed. Elsevier, N.Y. (1993).

Nucleic acids, vectors, capsids, polypeptides, and the like can be analyzed and quantified by any of a number of general means well known to those of skill in the art. These include, *e.g.*, analytical biochemical methods such as NMR, spectrophotometry, radiography, electrophoresis, capillary electrophoresis, high performance liquid chromatography (HPLC), thin layer chromatography (TLC), and hyperdiffusion chromatography, various immunological methods, *e.g.*, fluid or gel precipitin reactions, immunodiffusion, immunoelectrophoresis, radioimmunoassays (RIAs), enzyme-linked immunosorbent assays (ELISAs), immuno-fluorescent assays, Southern analysis, Northern analysis, dot-blot analysis, gel electrophoresis (*e.g.*, SDS-PAGE), RT-PCR, quantitative PCR, other nucleic acid or target or signal amplification methods, radiolabeling, scintillation counting, and affinity chromatography.

Oligonucleotide primers may be used to amplify nucleic acid fragments encoding taste receptor ligand-binding regions. The nucleic acids described herein can also be cloned or measured quantitatively using amplification techniques. Amplification methods are also well known in the art, and include, *e.g.*, polymerase chain reaction, PCR (*PCR Protocols, a Guide to Methods and Applications*, ed. Innis. Academic Press, N.Y. (1990) and *PCR Strategies*, ed. Innis, Academic Press, Inc., N.Y. (1995), ligase chain reaction (LCR) (*see, e.g.*, Wu, *Genomics* 4:560 (1989); Landegren, *Science* 241:1077, (1988); Barringer, *Gene* 89:117 (1990)); transcription amplification (*see, e.g.*, Kwoh, *Proc. Natl. Acad. Sci. USA* 86:1173 (1989)); and, self-sustained sequence replication (*see, e.g.*, Guatelli, *Proc. Natl. Acad. Sci. USA* 87:1874 (1990)); Q Beta replicase amplification (*see, e.g.*, Smith, *J. Clin. Microbiol.* 35:1477-1491 (1997));

automated Q-beta replicase amplification assay (*see, e.g.*, Burg, *Mol. Cell. Probes* 10:257-271 (1996)) and other RNA polymerase mediated techniques (*e.g.*, NASBA, Cangene, Mississauga, Ontario); *see also* Berger, *Methods Enzymol.* 152:307-316 (1987); Sambrook; Ausubel; U.S. Patent Nos. 4,683,195 and 4,683,202; Sooknanan, *Biotechnology* 13:563-564 (1995). The primers can be designed to retain the original sequence of the "donor" 7-membrane receptor. Alternatively, the primers can encode amino acid residues that are conservative substitutions (*e.g.*, hydrophobic for hydrophobic residue, see above discussion) or functionally benign substitutions (*e.g.*, do not prevent plasma membrane insertion, cause cleavage by peptidase, cause abnormal folding of receptor, and the like). Once amplified, the nucleic acids, either individually or as libraries, may be cloned according to methods known in the art, if desired, into any of a variety of vectors using routine molecular biological methods; methods for cloning *in vitro* amplified nucleic acids are described, *e.g.*, U.S. Pat. No. 5,426,039.

The primer pairs may be designed to selectively amplify ligand-binding regions of the T1R family members. These regions may vary for different ligands or tastants. Thus, what may be a minimal binding region for one tastant, may be too limiting for a second tastant. Accordingly, ligand-binding regions of different sizes comprising different extracellular domain structures may be amplified.

Paradigms to design degenerate primer pairs are well known in the art. For example, a COnsensus-DEgenerate Hybrid Oligonucleotide Primer (CODEHOP) strategy computer program is accessible as http://blocks.fhcrc.org/codehop.html, and is directly linked from the BlockMaker multiple sequence alignment site for hybrid primer prediction beginning with a set of related protein sequences, as known taste receptor ligand-binding regions (*see, e.g.*, Rose, *Nucleic Acids Res.* 26:1628-1635 (1998); Singh, *Biotechniques* 24:318-319 (1998)).

Means to synthesize oligonucleotide primer pairs are well known in the art. "Natural" base pairs or synthetic base pairs can be used. For example, use of artificial nucleobases offers a versatile approach to manipulate primer sequence and generate a more complex mixture of amplification products. Various families of artificial nucleobases are capable of assuming multiple hydrogen bonding orientations through internal bond rotations to provide a means for degenerate molecular recognition. Incorporation of these analogs into a single position of a PCR primer allows for

generation of a complex library of amplification products. *See, e.g.*, Hoops, *Nucleic Acids Res.* 25:4866-4871 (1997). Nonpolar molecules can also be used to mimic the shape of natural DNA bases. A non-hydrogen-bonding shape mimic for adenine can replicate efficiently and selectively against a nonpolar shape mimic for thymine (*see, e.g.*, Morales, *Nat. Struct. Biol.* 5:950-954 (1998)). For example, two degenerate bases can be the pyrimidine base 6H, 8H-3,4-dihydropyrimido[4,5-c][1,2]oxazin-7-one or the purine base N6-methoxy-2,6-diaminopurine (*see, e.g.*, Hill, *Proc. Natl. Acad. Sci. USA* 95:4258-4263 (1998)). Exemplary degenerate primers of the invention incorporate the nucleobase analog 5'-Dimethoxytrityl-N-benzoyl-2'-deoxy-Cytidine,3'-[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidite (the term "P" in the sequences, see above). This pyrimidine analog hydrogen bonds with purines, including A and G residues.

Polymorphic variants, alleles, and interspecies homologs that are substantially identical to a taste receptor disclosed herein can be isolated using the nucleic acid probes described above. Alternatively, expression libraries can be used to clone T1R polypeptides and polymorphic variants, alleles, and interspecies homologs thereof, by detecting expressed homologs immunologically with antisera or purified antibodies made against a T1R polypeptide, which also recognize and selectively bind to the T1R homolog.

Nucleic acids that encode ligand-binding regions of taste receptors may be generated by amplification (*e.g.*, PCR) of appropriate nucleic acid sequences using degenerate primer pairs. The amplified nucleic acid can be genomic DNA from any cell or tissue or mRNA or cDNA derived from taste receptor-expressing cells.

In one embodiment, hybrid protein-coding sequences comprising nucleic acids encoding T1Rs fused to translocation sequences may be constructed. Also provided are hybrid T1Rs comprising the translocation motifs and tastant-binding domains of other families of chemosensory receptors, particularly taste receptors. These nucleic acid sequences can be operably linked to transcriptional or translational control elements, *e.g.*, transcription and translation initiation sequences, promoters and enhancers, transcription and translation terminators, polyadenylation sequences, and other sequences useful for transcribing DNA into RNA. In constitutive of recombinant expression cassettes, vectors, and transgenics, a promoter fragment can be employed to direct expression of the desired nucleic acid in all desired cells or tissues.

In another embodiment, fusion proteins may include C-terminal or N-terminal translocation sequences. Further, fusion proteins can comprise additional elements, *e.g.*, for protein detection, purification, or other applications. Detection and purification facilitating domains include, *e.g.*, metal chelating peptides such as polyhistidine tracts, histidine-tryptophan modules, or other domains that allow purification on immobilized metals; maltose binding protein; protein A domains that allow purification on immobilized immunoglobulin; or the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, WA).

The inclusion of a cleavable linker sequences such as Factor Xa (*see, e.g.*, Ottavi, *Biochimie* 80:289-293 (1998)), subtilisin protease recognition motif (*see, e.g.*, Polyak, *Protein Eng.* 10:615-619 (1997)); enterokinase (Invitrogen, San Diego, CA), and the like, between the translocation domain (for efficient plasma membrane expression) and the rest of the newly translated polypeptide may be useful to facilitate purification. For example, one construct can include a polypeptide encoding a nucleic acid sequence linked to six histidine residues followed by a thioredoxin, an enterokinase cleavage site (*see, e.g.*, Williams, *Biochemistry* 34:1787-1797 (1995)), and an C-terminal translocation domain. The histidine residues facilitate detection and purification while the enterokinase cleavage site provides a means for purifying the desired protein(s) from the remainder of the fusion protein. Technology pertaining to vectors encoding fusion proteins and application of fusion proteins are well described in the scientific and patent literature, *see, e.g.*, Kroll, *DNA Cell. Biol.* 12:441-53 (1993).

Expression vectors, either as individual expression vectors or as libraries of expression vectors, comprising the ligand-binding domain encoding sequences may be introduced into a genome or into the cytoplasm or a nucleus of a cell and expressed by a variety of conventional techniques, well described in the scientific and patent literature. *See, e.g.*, Roberts, *Nature* 328:731 (1987); Berger *supra*; Schneider, *Protein Expr. Purif.* 6435:10 (1995); Sambrook; Tijssen; Ausubel. Product information from manufacturers of biological reagents and experimental equipment also provide information regarding known biological methods. The vectors can be isolated from natural sources, obtained from such sources as ATCC or GenBank libraries, or prepared by synthetic or recombinant methods.

The nucleic acids can be expressed using expression cassettes, vectors or viruses which are stably or transiently expressed in cells (*e.g.*, episomal expression

systems). Selection markers can be incorporated into expression cassettes and vectors to confer a selectable phenotype on transformed cells and sequences. For example, selection markers can code for episomal maintenance and replication such that integration into the host genome is not required. For example, the marker may encode antibiotic resistance (*e.g.*, chloramphenicol, kanamycin, G418, blasticidin, hygromycin) or herbicide resistance (*e.g.*, chlorosulfuron or Basta) to permit selection of those cells transformed with the desired DNA sequences (*see, e.g.*, Blondelet-Rouault, *Gene* 190:315-317 (1997); Aubrecht, *J. Pharmacol. Exp. Ther.* 281:992-997 (1997)). Because selectable marker genes conferring resistance to substrates like neomycin or hygromycin can only be utilized in tissue culture, chemoresistance genes are also used as selectable markers ***in vitro*** and ***in vivo***.

A chimeric nucleic acid sequence may encode a T1R ligand-binding domain within any 7-transmembrane polypeptide. Because 7-transmembrane receptor polypeptides have similar primary sequences and secondary and tertiary structures, structural domains (*e.g.*, extracellular domain, TM domains, cytoplasmic domain, *etc.*) can be readily identified by sequence analysis. For example, homology modeling, Fourier analysis and helical periodicity detection can identify and characterize the seven domains with a 7-transmembrane receptor sequence. Fast Fourier Transform (FFT) algorithms can be used to assess the dominant periods that characterize profiles of the hydrophobicity and variability of analyzed sequences. Periodicity detection enhancement and alpha helical periodicity index can be done as by, *e.g.*, Donnelly, *Protein Sci.* 2:55-70 (1993). Other alignment and modeling algorithms are well known in the art, *see, e.g.*, Peitsch, *Receptors Channels* 4:161-164 (1996); Kyte & Doolittle, *J. Med. Bio.*, 157:105-132 (1982); Cronet, *Protein Eng.* 6:59-64 (1993).

The present invention also includes not only the DNA and proteins having the specified nucleic and amino acid sequences, but also DNA fragments, particularly fragments of, *e.g.*, 40, 60, 80, 100, 150, 200, or 250 nucleotides, or more, as well as protein fragments of, *e.g.*, 10, 20, 30, 50, 70, 100, or 150 amino acids, or more. Optionally, the nucleic acid fragments can encode an antigenic polypeptide, which is capable of binding to an antibody raised against a T1R family member. Further, a protein fragment of the invention can optionally be an antigenic fragment, which is capable of binding to an antibody raised against a T1R family member.

Also contemplated are chimeric proteins, comprising at least 10, 20, 30, 50, 70, 100, or 150 amino acids, or more, of one of at least one of the T1R polypeptides described herein, coupled to additional amino acids representing all or part of another GPCR, preferably a member of the 7 transmembrane superfamily. These chimeras can be made from the instant receptors and another GPCR, or they can be made by combining two or more of the present T1R receptors. In one embodiment, one portion of the chimera corresponds to or is derived from the extracellular domain of a T1R polypeptide of the invention. In another embodiment, one portion of the chimera corresponds to, or is derived from the extracellular domain and one or more of the transmembrane domains of a T1R polypeptide described herein, and the remaining portion or portions can come from another GPCR. Chimeric receptors are well known in the art, and the techniques for creating them and the selection and boundaries of domains or fragments of G protein-coupled receptors for incorporation therein are also well known. Thus, this knowledge of those skilled in the art can readily be used to create such chimeric receptors. The use of such chimeric receptors can provide, for example, a taste selectivity characteristic of one of the receptors specifically disclosed herein, coupled with the signal transduction characteristics of another receptor, such as a well known receptor used in prior art assay systems.

As noted above, such chimeras, analogous to the native T1R receptor, or native T1R receptor combination or association will bind to and/or be activated by molecules that normally affect sweet taste or umami taste. Functional chimeric T1R receptors or receptor combinations are molecules which when expressed alone or in combination with other T1Rs or other GPCRs (which may themselves be chimeric) bind to or which are activated by taste stimuli, particularly sweet (T1R2/3) or umami taste stimuli (T1R1/3). Molecules that elicit sweet taste include natural and artificial sweeteners such as sucrose, aspartame, xylitol, cyclamate, et al., Molecules that elicit umami taste include glutamate and glutamate analogs and other compounds that bind to native T1R1 and/or T1R3, such as 5'-nucleotides.

For example, a domain such as a ligand-binding domain, an extracellular domain, a transmembrane domain, a transmembrane domain, a cytoplasmic domain, an N-terminal domain, a C-terminal domain, or any combination thereof, can be covalently linked to a heterologous protein. For instance, an T1R extracellular domain can be linked to a heterologous GPCR transmembrane domain, or a heterologous

GPCR extracellular domain can be linked to a T1R transmembrane domain. Other heterologous proteins of choice can be used; *e.g.*, green fluorescent protein.

Also within the scope of the invention are host cells for expressing the T1Rs, fragments, chimeras or variants of the invention. To obtain high levels of expression of a cloned gene or nucleic acid, such as cDNAs encoding the T1Rs, fragments, or variants of the invention, one of skill typically subclones the nucleic acid sequence of interest into an expression vector that contains a strong promoter to direct transcription, a transcription/translation terminator, and if for a nucleic acid encoding a protein, a ribosome binding site for translational initiation. Suitable bacterial promoters are well known in the art and described, *e.g.*, in Sambrook *et al.* However, bacterial or eukaryotic expression systems can be used.

Any of the well-known procedures for introducing foreign nucleotide sequences into host cells may be used. These include the use of calcium phosphate transfection, polybrene, protoplast fusion, electroporation, liposomes, microinjection, plasma vectors, viral vectors and any of the other well known methods for introducing cloned genomic DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell (*see, e.g.*, Sambrook *et al.*) It is only necessary that the particular genetic engineering procedure used be capable of successfully introducing at lest one nucleic acid molecule into the host cell capable of expressing the T1R, fragment, or variant of interest.

After the expression vector is introduced into the cells, the transfected cells are cultured under conditions favoring expression of the receptor, fragment, or variant of interest, which is then recovered from the culture using standard techniques. Examples of such techniques are well known in the art. *See, e.g.*, WO 00/06593, which is incorporated by reference in a manner consistent with this disclosure.

Detection of T1R polypeptides

In addition to the detection of T1R genes and gene expression using nucleic acid hybridization technology, one can also use immunoassays to detect T1Rs, *e.g.*, to identify taste receptor cells, and variants of T1R family members. Immunoassays can be used to qualitatively or quantitatively analyze the T1Rs. A general overview of the applicable technology can be found in Harlow & Lane, *Antibodies: A Laboratory Manual* (1988).

1. Antibodies to T1R family members

Methods of producing polyclonal and monoclonal antibodies that react specifically with a T1R family member are known to those of skill in the art (*see, e.g.*, Coligan, *Current Protocols in Immunology* (1991); Harlow & Lane, supra; Goding, *Monoclonal Antibodies: Principles and Practice* (2d ed. 1986); and Kohler & Milstein, *Nature*, 256:495-497 (1975)). Such techniques include antibody preparation by selection of antibodies from libraries of recombinant antibodies in phage or similar vectors, as well as preparation of polyclonal and monoclonal antibodies by immunizing rabbits or mice (*see, e.g.*, Huse *et al.*, *Science*, 246:1275-1281 (1989); Ward *et al.*, *Nature*, 341:544-546 (1989)).

A number of T1R-comprising immunogens may be used to produce antibodies specifically reactive with a T1R family member. For example, a recombinant T1R polypeptide, or an antigenic fragment thereof, can be isolated as described herein. Suitable antigenic regions include, *e.g.*, the consensus sequences that are used to identify members of the T1R family. Recombinant proteins can be expressed in eukaryotic or prokaryotic cells as described above, and purified as generally described above. Recombinant protein is the preferred immunogen for the production of monoclonal or polyclonal antibodies. Alternatively, a synthetic peptide derived from the sequences disclosed herein and conjugated to a carrier protein can be used an immunogen. Naturally occurring protein may also be used either in pure or impure form. The product is then injected into an animal capable of producing antibodies. Either monoclonal or polyclonal antibodies may be generated, for subsequent use in immunoassays to measure the protein.

Methods of production of polyclonal antibodies are known to those of skill in the art. For example, an inbred strain of mice (*e.g.*, BALB/C mice) or rabbits is immunized with the protein using a standard adjuvant, such as Freund's adjuvant, and a standard immunization protocol. The animal's immune response to the immunogen preparation is monitored by taking test bleeds and determining the titer of reactivity to the T1R. When appropriately high titers of antibody to the immunogen are obtained, blood is collected from the animal and antisera are prepared. Further fractionation of the antisera to enrich for antibodies reactive to the protein can be done if desired (*see* Harlow & Lane, *supra*).

Monoclonal antibodies may be obtained by various techniques familiar to those skilled in the art. Briefly, spleen cells from an animal immunized with a desired antigen may be immortalized, commonly by fusion with a myeloma cell (*see* Kohler & Milstein, *Eur. J. Immunol.*, 6:511-519 (1976)). Alternative methods of immortalization include transformation with Epstein Barr Virus, oncogenes, or retroviruses, or other methods well known in the art. Colonies arising from single immortalized cells are screened for production of antibodies of the desired specificity and affinity for the antigen, and yield of the monoclonal antibodies produced by such cells may be enhanced by various techniques, including injection into the peritoneal cavity of a vertebrate host. Alternatively, one may isolate DNA sequences which encode a monoclonal antibody or a binding fragment thereof by screening a DNA library from human B cells according to the general protocol outlined by Huse *et al.*, *Science*, 246:1275-1281 (1989).

Monoclonal antibodies and polyclonal sera are collected and titered against the immunogen protein in an immunoassay, for example, a solid phase immunoassay with the immunogen immobilized on a solid support. Typically, polyclonal antisera with a titer of 104 or greater are selected and tested for their cross reactivity against non-T1R polypeptides, or even other T1R family members or other related proteins from other organisms, using a competitive binding immunoassay. Specific polyclonal antisera and monoclonal antibodies will usually bind with a Kd of at least about 0.1 mM, more usually at least about 1 pM, optionally at least about 0.1 pM or better, and optionally 0.01 pM or better.

Once T1R family member specific antibodies are available, individual T1R proteins and protein fragments can be detected by a variety of immunoassay methods. For a review of immunological and immunoassay procedures, *see Basic and Clinical Immunology* (Stites & Terr eds., 7th ed. 1991). Moreover, the immunoassays of the present invention can be performed in any of several configurations, which are reviewed extensively in Enzyme Immunoassay (Maggio, ed., 1980); and Harlow & Lane, *supra*.

2. Immunological binding assays

T1R proteins, fragments, and variants can be detected and/or quantified using any of a number of well-recognized immunological binding assays (*see, e.g.*, U.S.

Patents 4,366,241; 4,376,110; 4,517,288; and 4,837,168). For a review of the general immunoassays, *see also Methods in Cell Biology: Antibodies in Cell Biology*, volume 37 (Asai, ed. 1993); *Basic and Clinical Immunology* (Stites & Terr, eds., 7th ed. 1991). Immunological binding assays (or immunoassays) typically use an antibody that specifically binds to a protein or antigen of choice (in this case a T1R family member or an antigenic subsequence thereof). The antibody (*e.g.*, anti-T1R) may be produced by any of a number of means well known to those of skill in the art and as described above.

Immunoassays also often use a labeling agent to specifically bind to and label the complex formed by the antibody and antigen. The labeling agent may itself be one of the moieties comprising the antibody/antigen complex. Thus, the labeling agent may be a labeled T1R polypeptide or a labeled anti-T1R antibody. Alternatively, the labeling agent may be a third moiety, such a secondary antibody that specifically binds to the antibody/T1R complex (a secondary antibody is typically specific to antibodies of the species from which the first antibody is derived). Other proteins capable of specifically binding immunoglobulin constant regions, such as protein A or protein G may also be used as the label agent. These proteins exhibit a strong non-immunogenic reactivity with immunoglobulin constant regions from a variety of species (*see, e.g.*, Kronval *et al.*, *J. Immunol.*, 111:1401-1406 (1973); *Akerstrom et al.*, *J. Immunol.*, 135:2589-2542 (1985)). The labeling agent can be modified with a detectable moiety, such as biotin, to which another molecule can specifically bind, such as streptavidin. A variety of detectable moieties are well known to those skilled in the art.

Throughout the assays, incubation and/or washing steps may be required after each combination of reagents. Incubation steps can vary from about 5 seconds to several hours, optionally from about 5 minutes to about 24 hours. However, the incubation time will depend upon the assay format, antigen, volume of solution, concentrations, and the like. Usually, the assays will be carried out at ambient temperature, although they can be conducted over a range of temperatures, such as 10°C to 40°C.

A. Non-competitive assay formats

Immunoassays for detecting a T1R polypeptide in a sample may be either competitive or noncompetitive. Noncompetitive immunoassays are assays in which the

amount of antigen is directly measured. In one preferred "sandwich" assay, for example, the anti-T1R antibodies can be bound directly to a solid substrate on which they are immobilized. These immobilized antibodies then capture the T1R polypeptide present in the test sample. The T1R polypeptide is thus immobilized is then bound by a labeling agent, such as a second T1R antibody bearing a label. Alternatively, the second antibody may lack a label, but it may, in turn, be bound by a labeled third antibody specific to antibodies of the species from which the second antibody is derived. The second or third antibody is typically modified with a detectable moiety, such as biotin, to which another molecule specifically binds, *e.g.*, streptavidin, to provide a detectable moiety.

B. Competitive assay formats

In competitive assays, the amount of T1R polypeptide present in the sample is measured indirectly by measuring the amount of a known, added (exogenous) T1R polypeptide displaced (competed away) from an anti-T1R antibody by the unknown T1R polypeptide present in a sample. In one competitive assay, a known amount of T1R polypeptide is added to a sample and the sample is then contacted with an antibody that specifically binds to the T1R. The amount of exogenous T1R polypeptide bound to the antibody is inversely proportional to the concentration of T1R polypeptide present in the sample. In a particularly preferred embodiment, the antibody is immobilized on a solid substrate. The amount of T1R polypeptide bound to the antibody may be determined either by measuring the amount of T1R polypeptide present in a T1R/antibody complex, or alternatively by measuring the amount of remaining uncomplexed protein. The amount of T1R polypeptide may be detected by providing a labeled T1R molecule.

A hapten inhibition assay is another preferred competitive assay. In this assay the known T1R polypeptide is immobilized on a solid substrate. A known amount of anti-T1R antibody is added to the sample, and the sample is then contacted with the immobilized T1R. The amount of anti-T1R antibody bound to the known immobilized T1R polypeptide is inversely proportional to the amount of T1R polypeptide present in the sample. Again, the amount of immobilized antibody may be detected by detecting either the immobilized fraction of antibody or the fraction of the antibody that remains in solution. Detection may be direct where the antibody is labeled or indirect by the

subsequent addition of a labeled moiety that specifically binds to the antibody as described above.

C. Cross-reactivity determinations

Immunoassays in the competitive binding format can also be used for cross-reactivity determinations. For example, a protein at least partially encoded by the nucleic acid sequences disclosed herein can be immobilized to a solid support. Proteins (*e.g.*, T1R polypeptides and homologs) are added to the assay that compete for binding of the antisera to the immobilized antigen. The ability of the added proteins to compete for binding of the antisera to the immobilized protein is compared to the ability of the T1R polypeptide encoded by the nucleic acid sequences disclosed herein to compete with itself. The percent cross-reactivity for the above proteins is calculated, using standard calculations. Those antisera with less than 10% cross-reactivity with each of the added proteins listed above are selected and pooled. The cross-reacting antibodies are optionally removed from the pooled antisera by immunoabsorption with the added considered proteins, *e.g.*, distantly related homologs. In addition, peptides comprising amino acid sequences representing conserved motifs that are used to identify members of the T1R family can be used in cross-reactivity determinations.

The immunoabsorbed and pooled antisera are then used in a competitive binding immunoassay as described above to compare a second protein, thought to be perhaps an allele or polymorphic variant of a T1R family member, to the immunogen protein (*i.e.*, T1R polypeptide encoded by the nucleic acid sequences disclosed herein). In order to make this comparison, the two proteins are each assayed at a wide range of concentrations and the amount of each protein required to inhibit 50% of the binding of the antisera to the immobilized protein is determined. If the amount of the second protein required to inhibit 50% of binding is less than 10 times the amount of the protein encoded by nucleic acid sequences disclosed herein required to inhibit 50% of binding, then the second protein is said to specifically bind to the polyclonal antibodies generated to a T1R immunogen.

Antibodies raised against T1R conserved motifs can also be used to prepare antibodies that specifically bind only to GPCRs of the T1R family, but not to GPCRs from other families.

Polyclonal antibodies that specifically bind to a particular member of the T1R family can be made by subtracting out cross-reactive antibodies using other T1R family members. Species-specific polyclonal antibodies can be made in a similar way. For example, antibodies specific to human T1R1 can be made by, subtracting out antibodies that are cross-reactive with orthologous sequences, *e.g.*, rat T1R1 or mouse T1R1.

D. Other assay formats

Western blot (immunoblot) analysis is used to detect and quantify the presence of T1R polypeptide in the sample. The technique generally comprises separating sample proteins by gel electrophoresis on the basis of molecular weight, transferring the separated proteins to a suitable solid support, (such as a nitrocellulose filter, a nylon filter, or derivatized nylon filter), and incubating the sample with the antibodies that specifically bind the T1R polypeptide. The anti-T1R polypeptide antibodies specifically bind to the T1R polypeptide on the solid support. These antibodies may be directly labeled or alternatively may be subsequently detected using labeled antibodies (*e.g.*, labeled sheep anti-mouse antibodies) that specifically bind to the anti-T1R antibodies.

Other, assay formats include liposome immunoassays (LIA), which use liposomes designed to bind specific molecules (*e.g.*, antibodies) and release encapsulated reagents or markers. The released chemicals are then detected according to standard techniques (*see* Monroe *et al.*, *Amer. Clin. Prod. Rev.*, 5:34-41 (1986)).

E. Reduction of non-specific binding

One of skill in the art will appreciate that it is often desirable to minimize non-specific binding in immunoassays. Particularly, where the assay involves an antigen or antibody immobilized on a solid substrate it is desirable to minimize the amount of non-specific binding to the substrate. Means of reducing such non-specific binding are well known to those of skill in the art. Typically, this technique involves coating the substrate with a proteinaceous composition. In particular, protein compositions such as bovine serum albumin (BSA), nonfat powdered milk, and gelatin are widely used with powdered milk being most preferred.

F. Labels

The particular label or detectable group used in the assay is not a critical aspect of the invention, as long as it does not significantly interfere with the specific binding of the antibody used in the assay. The detectable group can be any material having a detectable physical or chemical property. Such detectable labels have been well developed in the field of immunoassays and, in general, most any label useful in such methods can be applied to the present invention. Thus, a label is any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical, or chemical means. Useful labels in the present invention include magnetic beads (*e.g.*, DYNABEADSTM), fluorescent dyes (*e.g.*, fluorescein isothiocyanate, Texas red, rhodamine, and the like), radiolabels (*e.g.*, ^{3}H, ^{125}I, ^{14}C, ^{35}S), enzymes (*e.g.*, horseradish peroxidase, alkaline phosphates and others commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic beads (*e.g.*, polystyrene, polypropylene, latex, *etc.*).

The label may be coupled directly or indirectly to the desired component of the assay according to methods well known in the art. As indicated above, a wide variety of labels may be used, with the choice of label depending on sensitivity required, ease of conjugation with the compound, stability requirements, available instrumentation, and disposal provisions.

Non-radioactive labels are often attached by indirect means. Generally, a ligand molecule (*e.g.*, biotin) is covalently bound to the molecule. The ligand then binds to another molecules (*e.g.*, streptavidin) molecule, which is either inherently detectable or covalently bound to a signal system, such as a detectable enzyme, a fluorescent compound, or a chemiluminescent compound. The ligands and their targets can be used in any suitable combination with antibodies that recognize a T1R polypeptide, or secondary antibodies that recognize anti-T1R.

The molecules can also be conjugated directly to signal generating compounds, *e.g.*, by conjugation with an enzyme or fluorophore. Enzymes of interest as labels will primarily be hydrolases, particularly phosphatases, esterases and glycosidases, or oxidotases, particularly peroxidases. Fluorescent compounds include fluorescein and its derivatives, rhodamine and its derivatives, dansyl, umbelliferone, *etc.* Chemiluminescent compounds include luciferin, and 2,3-dihydrophthalazinediones, *e.g.*, luminol. For a review of various labeling or signal producing systems that may be used, *see* U.S. Patent No. 4,391,904.

Means of detecting labels are well known to those of skill in the art. Thus, for example, where the label is a radioactive label, means for detection include a scintillation counter or photographic film as in autoradiography. Where the label is a fluorescent label, it may be detected by exciting the fluorochrome with the appropriate wavelength of light and detecting the resulting fluorescence. The fluorescence may be detected visually, by means of photographic film, by the use of electronic detectors such as charge-coupled devices (CCDs) or photomultipliers and the like. Similarly, enzymatic labels may be detected by providing the appropriate substrates for the enzyme and detecting the resulting reaction product. Finally simple colorimetric labels may be detected simply by observing the color associated with the label. Thus, in various dipstick assays, conjugated gold often appears pink, while various conjugated beads appear the color of the bead.

Some assay formats do not require the use of labeled components. For instance, agglutination assays can be used to detect the presence of the target antibodies. In this case, antigen-coated particles are agglutinated by samples comprising the target antibodies. In this format, none of the components need be labeled and the presence of the target antibody is detected by simple visual inspection.

Detection of Modulators

Compositions and methods for determining whether a test compound specifically binds to a T1R receptor of the invention, both *in vitro* and *in vivo*, are described below. Many aspects of cell physiology can be monitored to assess the effect of ligand binding to a T1R polypeptide of the invention. These assays may be performed on intact cells expressing a chemosensory receptor, on permeabilized cells, or on membrane fractions produced by standard methods or in vitro de novo synthesized proteins.

In vivo, taste receptors bind tastants and initiate the transduction of chemical stimuli into electrical signals. An activated or inhibited G protein will in turn alter the properties of target enzymes, channels, and other effector proteins. Some examples are the activation of cGMP phosphodiesterase by transducin in the visual system, adenylate cyclase by the stimulatory G protein, phospholipase C by Gq and other cognate G proteins, and modulation of diverse channels by Gi and other G proteins. Downstream

consequences can also be examined such as generation of diacyl glycerol and IP3 by phospholipase C, and in turn, for calcium mobilization by IP3.

The T1R proteins or polypeptides of the assay will preferably be selected from a polypeptide having the T1R polypeptide sequence selected from those disclosed in Example 1, or fragments or conservatively modified variants thereof. Optionally, the fragments and variants can be antigenic fragments and variants which bind to an anti-T1R antibody. Optionally, the fragments and variants can bind to or are activated by sweeteners or umami tastants.

Alternatively, the T1R proteins or polypeptides of the assay can be derived from a eukaryotic host cell and can include an amino acid subsequence having amino acid sequence identity to the T1R polypeptides disclosed in Example 1, or fragments or conservatively modified variants thereof. Generally, the amino acid sequence identity will be at least 35 to 50%, or optionally 75%, 85%, 90%, 95%, 96%, 97%, 98%, or 99%. Optionally, the T1R proteins or polypeptides of the assays can comprise a domain of a T1R protein, such as an extracellular domain, transmembrane region, transmembrane domain, cytoplasmic domain, ligand-binding domain, and the like. Further, as described above, the T1R protein or a domain thereof can be covalently linked to a heterologous protein to create a chimeric protein used in the assays described herein.

Modulators of T1R receptor activity are tested using T1R proteins or polypeptides as described above, either recombinant or naturally occurring. The T1R proteins or polypeptides can be isolated, co-expressed in a cell, co-expressed in a membrane derived from a cell, co-expressed in tissue or in an animal, either recombinant or naturally occurring. For example, tongue slices, dissociated cells from a tongue, transformed cells, or membranes can be used. Modulation can be tested using one of the *in vitro* or *in vivo* assays described herein.

For example, as disclosed in the experiment examples infra, it has been discovered that certain 5' nucleotides, e.g., 5' IMP or 5' GMP, enhance the activity of L-glutamate to activate the umami taste receptor, or block the activation of the umami taste receptor by umami taste stimuli such as L-glutamate and L-aspartate.

1. *In vitro* binding assays

Taste transduction can also be examined *in vitro* with soluble or solid state reactions, using the T1R polypeptides of the invention. In a particular embodiment, T1R ligand-binding domains can be used *in vitro* in soluble or solid state reactions to assay for ligand binding.

For instance, the T1R N-terminal domain is predicted to be involved in ligand binding. More particularly, the T1Rs belong to a GPCR sub-family that is characterized by large, approximately 600 amino acid, extracellular N-terminal segments. These N-terminal segments are thought to form the ligand-binding domains, and are therefore useful in biochemical assays to identify T1R agonists and antagonists. It is possible that the ligand-binding domain may be formed by additional portions of the extracellular domain, such as the extracellular loops of the transmembrane domain.

In vitro binding assays have been used with other GPCRs that are related to the T1Rs, such as the metabotropic glutamate receptors (*see, e.g.*, Han and Hampson, *J. Biol. Chem.* 274:10008-10013 (1999)). These assays might involve displacing a radioactively or fluorescently labeled ligand, measuring changes in intrinsic fluorescence or changes in proteolytic susceptibility, *etc.*

Ligand binding to a hetero-multimeric complex of T1R polypeptides of the invention can be tested in solution, in a bilayer membrane, optionally attached to a solid phase, in a lipid monolayer, or in vesicles. Binding of a modulator can be tested using, *e.g.*, changes in spectroscopic characteristics (*e.g.*, fluorescence, absorbence, refractive index) hydrodynamic (*e.g.*, shape), chromatographic, or solubility properties.

[0207] In another embodiment of the invention, a GTPγ^{35}S assay may be used. As described above, upon activation of a GPCR, the Gα subunit of the G protein complex is stimulated to exchange bound GDP for GTP. Ligand-mediated stimulation of G protein exchange activity can be measured in a biochemical assay measuring the binding of added radioactively labeled GTPγ^{35}S to the G protein in the presence of a putative ligand. Typically, membranes containing the chemosensory receptor of interest are mixed with a complex of G proteins. Potential inhibitors and/or activators and GTPγ^{35}S are added to the assay, and binding of GTPγ^{35}S to the G protein is measured. Binding can be measured by liquid scintillation counting or by any other means known in the art, including scintillation proximity assays (SPA). In other assays formats, fluorescently labeled GTPγS can be utilized.

2. Fluorescence Polarization Assays

In another embodiment, Fluorescence Polarization ("FP") based assays may be used to detect and monitor ligand binding. Fluorescence polarization is a versatile laboratory technique for measuring equilibrium binding, nucleic acid hybridization, and enzymatic activity. Fluorescence polarization assays are homogeneous in that they do not require a separation step such as centrifugation, filtration, chromatography, precipitation, or electrophoresis. These assays are done in real time, directly in solution and do not require an immobilized phase. Polarization values can be measured repeatedly and after the addition of reagents since measuring the polarization is rapid and does not destroy the sample. Generally, this technique can be used to measure polarization values of fluorophores from low picomolar to micromolar levels. This section describes how fluorescence polarization can be used in a simple and quantitative way to measure the binding of ligands to the T1R polypeptides of the invention.

When a fluorescently labeled molecule is excited with plane-polarized light, it emits light that has a degree of polarization that is inversely proportional to its molecular rotation. Large fluorescently labeled molecules remain relatively stationary during the excited state (4 nanoseconds in the case of fluorescein) and the polarization of the light remains relatively constant between excitation and emission. Small fluorescently labeled molecules rotate rapidly during the excited state and the polarization changes significantly between excitation and emission. Therefore, small molecules have low polarization values and large molecules have high polarization values. For example, a single-stranded fluorescein-labeled oligonucleotide has a relatively low polarization value but when it is hybridized to a complementary strand, it has a higher polarization value. When using FP to detect and monitor tastant-binding which may activate or inhibit the chemosensory receptors of the invention, fluorescence-labeled tastants or auto-fluorescent tastants may be used.

Fluorescence polarization (P) is defined as:

$$P = \frac{Int_{\Pi} - Int_{\perp}}{Int_{\Pi} + Int_{\perp}}$$

Where Π is the intensity of the emission light parallel to the excitation light plane and Int $\perp$ is the intensity of the emission light perpendicular to the excitation light plane. P, being a ratio of light intensities, is a dimensionless number. For example, the Beacon

® and Beacon 2000 ™ System may be used in connection with these assays. Such systems typically express polarization in millipolarization units (1 Polarization Unit =1000 mP Units).

The relationship between molecular rotation and size is described by the Perrin equation and the reader is referred to Jolley, M. E. (1991) in Journal of Analytical Toxicology, pp. 236-240, which gives a thorough explanation of this equation. Summarily, the Perrin equation states that polarization is directly proportional to the rotational relaxation time, the time that it takes a molecule to rotate through an angle of approximately 68.5° Rotational relaxation time is related to viscosity (η), absolute temperature (T), molecular volume (V), and the gas constant (R) by the following equation:

$$Rotational\,\mathrm{Re}laxationTime = \frac{3\eta V}{RT}$$

The rotational relaxation time is small (≈ 1 nanosecond) for small molecules (*e.g.* fluorescein) and large (≈ 100 nanoseconds) for large molecules (*e.g.* immunoglobulins). If viscosity and temperature are held constant, rotational relaxation time, and therefore polarization, is directly related to the molecular volume. Changes in molecular volume may be due to interactions with other molecules, dissociation, polymerization, degradation, hybridization, or conformational changes of the fluorescently labeled molecule. For example, fluorescence polarization has been used to measure enzymatic cleavage of large fluorescein labeled polymers by proteases, DNases, and RNases. It also has been used to measure equilibrium binding for protein/protein interactions, antibody/antigen binding, and protein/DNA binding.

A. <u>Solid state and soluble high throughput assays</u>

In yet another embodiment, the invention provides soluble assays using a hetero-oligomeric T1R polypeptide complex; or a cell or tissue co-expressing T1R polypeptides. Preferably, the cell will comprise a cell line that stably co-expresses a functional T1R1/T1R3 (umami) taste receptor or T1R2/T1R3 (sweet) taste receptor. In another embodiment, the invention provides solid phase based *in vitro* assays in a high throughput format, where the T1R polypeptides, or cell or tissue expressing the T1R polypeptides is attached to a solid phase substrate or a taste stimulating compound and

contacted with a T1R receptor, and binding detected using an appropriate tag or antibody raised against the T1R receptor.

In the high throughput assays of the invention, it is possible to screen up to several thousand different modulators or ligands in a single day. In particular, each well of a microtiter plate can be used to run a separate assay against a selected potential modulator, or, if concentration or incubation time effects are to be observed, every 5-10 wells can test a single modulator. Thus, a single standard microtiter plate can assay about 100 (*e.g.*, 96) modulators. If 1536 well plates are used, then a single plate can easily assay from about 1000 to about 1500 different compounds. It is also possible to assay multiple compounds in each plate well. It is possible to assay several different plates per day; assay screens for up to about 6,000-20,000 different compounds is possible using the integrated systems of the invention. More recently, microfluidic approaches to reagent manipulation have been developed.

The molecule of interest can be bound to the solid state component, directly or indirectly, via covalent or non-covalent linkage, *e.g.*, via a tag. The tag can be any of a variety of components. In general, a molecule which binds the tag (a tag binder) is fixed to a solid support, and the tagged molecule of interest (*e.g.*, the taste transduction molecule of interest) is attached to the solid support by interaction of the tag and the tag binder.

A number of tags and tag binders can be used, based upon known molecular interactions well described in the literature. For example, where a tag has a natural binder, for example, biotin, protein A, or protein G, it can be used in conjunction with appropriate tag binders (avidin, streptavidin, neutravidin, the Fc region of an immunoglobulin, *etc.*). Antibodies to molecules with natural binders such as biotin are also widely available and appropriate tag binders (*see*, SIGMA Immunochemicals 1998 catalogue SIGMA, St. Louis MO).

Similarly, any haptenic or antigenic compound can be used in combination with an appropriate antibody to form a tag/tag binder pair. Thousands of specific antibodies are commercially available and many additional antibodies are described in the literature. For example, in one common configuration, the tag is a first antibody and the tag binder is a second antibody which recognizes the first antibody. In addition to antibody-antigen interactions, receptor-ligand interactions are also appropriate as tag and tag-binder pairs. For example, agonists and antagonists of cell membrane receptors

(*e.g.*, cell receptor-ligand interactions such as transferrin, c-kit, viral receptor ligands, cytokine receptors, chemokine receptors, interleukin receptors, immunoglobulin receptors and antibodies, the cadherein family, the integrin family, the selectin family, and the like; *see*, *e.g.*, Pigott & Power, The Adhesion Molecule Facts Book I (1993)). Similarly, toxins and venoms, viral epitopes, hormones (*e.g.*, opiates, steroids, *etc.*), intracellular receptors (*e.g.*, which mediate the effects of various small ligands, including steroids, thyroid hormone, retinoids and vitamin D; peptides), drugs, lectins, sugars, nucleic acids (both linear and cyclic polymer configurations), oligosaccharides, proteins, phospholipids and antibodies can all interact with various cell receptors.

Synthetic polymers, such as polyurethanes, polyesters, polycarbonates, polyureas, polyamides, polyethyleneimines, polyarylene sulfides, polysiloxanes, polyimides, and polyacetates can also form an appropriate tag or tag binder. Many other tag/tag binder pairs are also useful in assay systems described herein, as would be apparent to one of skill upon review of this disclosure.

Common linkers such as peptides, polyethers, and the like can also serve as tags, and include polypeptide sequences, such as poly gly sequences of between about 5 and 200 amino acids. Such flexible linkers are known to persons of skill in the art. For example, poly(ethelyne glycol) linkers are available from Shearwater Polymers, Inc. Huntsville, Alabama. These linkers optionally have amide linkages, sulfhydryl linkages, or heterofunctional linkages.

Tag binders are fixed to solid substrates using any of a variety of methods currently available. Solid substrates are commonly derivatized or functionalized by exposing all or a portion of the substrate to a chemical reagent which fixes a chemical group to the surface which is reactive with a portion of the tag binder. For example, groups which are suitable for attachment to a longer chain portion would include amines, hydroxyl, thiol, and carboxyl groups. Aminoalkylsilanes and hydroxyalkylsilanes can be used to functionalize a variety of surfaces, such as glass surfaces. The constitutive of such solid phase biopolymer arrays is well described in the literature. *See, e.g.*, Merrifield, *J. Am. Chem. Soc.*, 85:2149-2154 (1963) (describing solid phase synthesis of, *e.g.*, peptides); Geysen *et al.*, *J. Immun. Meth.*, 102:259-274 (1987) (describing synthesis of solid phase components on pins); Frank & Doring, *Tetrahedron*, 44:60316040 (1988) (describing synthesis of various peptide sequences on cellulose disks); Fodor *et al., Science*, 251:767-777 (1991); *Sheldon et*

al., Clinical Chemistry, 39(4):718-719 (1993); and Kozal *et al., Nature Medicine*, 2(7):753759 (1996) (all describing arrays of biopolymers fixed to solid substrates). Non-chemical approaches for fixing tag binders to substrates include other common methods, such as heat, cross-linking by UV radiation, and the like.

3. Cell-based assays

In a preferred embodiment of treatment, a combination of T1R proteins or polypeptides are transiently or stably co-expressed in a eukaryotic cell either in unmodified forms or as chimeric, variant or truncated receptors with or preferably without a heterologous, chaperone sequence that facilitates its maturation and targeting through the secretory pathway. Such T1R polypeptides can be expressed in any eukaryotic cell, such as HEK-293 cells. Preferably, the cells comprise a functional G protein, *e.g.*, Gα15 or the chimeric G protein previously identified, or another G protein that is capable of coupling the chimeric receptor to an intracellular signaling pathway or to a signaling protein such as phospholipase C. Also, preferably a cell will be produced that stably co-expresses T1R1/T1R3 or T1R2/T1R3 as such cells have been found (as shown in the experimental examples) to exhibit enhanced responses to taste stimuli (relation to cells that transiently express the same T1R combination). Activation of T1R receptors in such cells can be detected using any standard method, such as by detecting changes in intracellular calcium by detecting Fluo-4 dependent fluorescence in the cell. Such an assay is the basis of the experimental findings presented in this application.

Activated GPCR receptors often are substrates for kinases that phosphorylate the C-terminal tail of the receptor (and possibly other sites as well). Thus, activators will promote the transfer of ^{32}P from radiolabeled ATP to the receptor, which can be assayed with a scintillation counter. The phosphorylation of the C-terminal tail will promote the binding of arrestin-like proteins and will interfere with the binding of G proteins. For a general review of GPCR signal transduction and methods of assaying signal transduction, *see, e.g., Methods in Enzymology*, vols. 237 and 238 (1994) and volume 96 (1983); Bourne *et al., Nature*, 10:349:117-27 (1991); Bourne *et al., Nature*, 348:125-32 (1990); Pitcher *et al., Annu. Rev. Biochem.*, 67:653-92 (1998).

T1R modulation may be assayed by comparing the response of T1R polypeptides treated with a putative T1R modulator to the response of an untreated

control sample or a sample containing a known "positive" control. Such putative T1R modulators can include molecules that either inhibit or activate T1R polypeptide activity. In one embodiment, control samples (untreated with activators or inhibitors) are assigned a relative T1R activity value of 100. Inhibition of a T1R polypeptide is achieved when the T1R activity value relative to the control is about 90%, optionally 50%, optionally 25-0%. Activation of a T1R polypeptide is achieved when the T1R activity value relative to the control is 110%, optionally 150%, 200-500%, or 1000-2000%.

Changes in ion flux may be assessed by determining changes in ionic polarization (*i.e.*, electrical potential) of the cell or membrane expressing a T1R polypeptide. One means to determine changes in cellular polarization is by measuring changes in current (thereby measuring changes in polarization) with voltage-clamp and patch-clamp techniques (*see, e.g.*, the "cell-attached" mode, the "inside-out" mode, and the "whole cell" mode, *e.g.*, Ackerman *et al.*, *New Engl. J Med.*, 336:1575-1595 (1997)). Whole cell currents are conveniently determined using the standard. Other known assays include: radiolabeled ion flux assays and fluorescence assays using voltage-sensitive dyes (*see, e.g.*, Vestergarrd-Bogind *et al.*, *J. Membrane Biol.*, 88:67-75 (1988); Gonzales & Tsien, *Chem. Biol.*, 4:269277 (1997); Daniel *et al.*, *J. Pharmacol. Meth.*, 25:185-193 (1991); Holevinsky *et al.*, *J. Membrane Biology*, 137:59-70 (1994)).

The effects of the test compounds upon the function of the polypeptides can be measured by examining any of the parameters described above. Any suitable physiological change that affects GPCR activity can be used to assess the influence of a test compound on the polypeptides of this invention. When the functional consequences are determined using intact cells or animals, one can also measure a variety of effects such as transmitter release, hormone release, transcriptional changes to both known and uncharacterized genetic markers (*e.g.*, northern blots), changes in cell metabolism such as cell growth or pH changes, and changes in intracellular second messengers such as Ca^{2+}, IP3, cGMP, or cAMP.

Preferred assays for GPCRs include cells that are loaded with ion or voltage sensitive dyes to report receptor activity. Assays for determining activity of such receptors can-also use known agonists and antagonists for other G protein-coupled receptors as controls to assess activity of tested compounds. In assays for identifying

modulatory compounds (*e.g.*, agonists, antagonists), changes in the level of ions in the cytoplasm or membrane voltage will be monitored using an ion sensitive or membrane voltage fluorescent indicator, respectively. Among the ion-sensitive indicators and voltage probes that may be employed are those disclosed in the Molecular Probes 1997 Catalog. For G protein-coupled receptors, promiscuous G proteins such as Gα15 and Gα16 can be used in the assay of choice (Wilkie *et al.*, *Proc. Nat'l Acad. Sci.*, 88:10049-10053 (1991)).

Receptor activation initiates subsequent intracellular events, *e.g.*, increases in second messengers. Activation of some G protein-coupled receptors stimulates the formation of inositol triphosphate (IP3) through phospholipase C-mediated hydrolysis of phosphatidylinositol (Berridge & Irvine, *Nature*, 312:315-21 (1984)). IP3 in turn stimulates the release of intracellular calcium ion stores. Thus, a change in cytoplasmic calcium ion levels, or a change in second messenger levels such as IP3 can be used to assess G protein-coupled receptor function. Cells expressing such G protein-coupled receptors may exhibit increased cytoplasmic calcium levels as a result of contribution from both calcium release from intracellular stores and extracellular calcium entry via plasma membrane ion channels.

In a preferred embodiment, T1R polypeptide activity is measured by stably or transiently co-expressing T1R genes, preferably stably, in a heterologous cell with a promiscuous G protein that links the receptor to a phospholipase C signal transduction pathway (*see* Offermanns & Simon, *J. Biol. Chem.*, 270:15175-15180 (1995)). In a preferred embodiment, the cell line is HEK-293 (which does not normally express T1R genes) and the promiscuous G protein is Gα15 (Offermanns & Simon, *supra*). Modulation of taste transduction is assayed by measuring changes in intracellular Ca^{2+} levels, which change in response to modulation of the T1R signal transduction pathway via administration of a molecule that associates with T1R polypeptides. Changes in Ca^{2+} levels are optionally measured using fluorescent Ca^{2+} indicator dyes and fluorometric imaging.

In another embodiment, phosphatidyl inositol (PI) hydrolysis can be analyzed according to U.S. Patent 5,436,128, herein incorporated by reference. Briefly, the assay involves labeling of cells with 3H-myoinositol for 48 or more hrs. The labeled cells are treated with a test compound for one hour. The treated cells are lysed and extracted in chloroform-methanol-water after which the inositol phosphates were

separated by ion exchange chromatography and quantified by scintillation counting. Fold stimulation is determined by calculating the ratio of cpm in the presence of agonist, to cpm in the presence of buffer control. Likewise, fold inhibition is determined by calculating the ratio of cpm in the presence of antagonist, to cpm in the presence of buffer control (which may or may not contain an agonist).

Other receptor assays can involve determining the level of intracellular cyclic nucleotides, *e.g.*, cAMP or cGMP. In cases where activation of the receptor results in a decrease in cyclic nucleotide levels, it may be preferable to expose the cells to agents that increase intracellular cyclic nucleotide levels, *e.g.*, forskolin, prior to adding a receptor-activating compound to the cells in the assay. In one embodiment, the changes in intracellular cAMP or cGMP can be measured using immunoassays. The method described in Offermanns & Simon, *J. Bio. Chem.*, 270:15175-15180 (1995), may be used to determine the level of cAMP. Also, the method described in Felley-Bosco *et al.*, *Am. J. Resp. Cell and Mol. Biol.*, 11:159-164 (1994), may be used to determine the level of cGMP. Further, an assay kit for measuring cAMP and/or cGMP is described in U.S. Patent 4,115,538, herein incorporated by reference.

In another embodiment, transcription levels can be measured to assess the effects of a test compound on signal transduction. A host cell containing T1R polypeptides of interest is contacted with a test compound for a sufficient time to effect any interactions, and then the level of gene expression is measured. The amount of time to effect such interactions may be empirically determined, such as by running a time course and measuring the level of transcription as a function of time. The amount of transcription may be measured by using any method known to those of skill in the art to be suitable. For example, mRNA expression of the protein of interest may be detected using northern blots or their polypeptide products may be identified using immunoassays. Alternatively, transcription based assays using reporter gene may be used as described in U.S. Patent 5,436,128, herein incorporated by reference. The reporter genes can be, *e.g.*, chloramphenicol acetyltransferase, luciferase, beta-galactosidase beta-lactamase and alkaline phosphatase. Furthermore, the protein of interest can be used as an indirect reporter via attachment to a second reporter such as green fluorescent protein (*see, e.g.*, Mistili & Spector, *Nature Biotechnology*, 15:961-964 (1997)).

The amount of transcription is then compared to the amount of transcription in either the same cell in the absence of the test compound, or it may be compared with the amount of transcription in a substantially identical cell that lacks the T1R polypeptide(s) of interest. A substantially identical cell may be derived from the same cells from which the recombinant cell was prepared but which had not been modified by introduction of heterologous DNA. Any difference in the amount of transcription indicates that the test compound has in some manner altered the activity of the T1R polypeptides of interest.

4. Transgenic non-human animals expressing chemosensory receptors

Non-human animals expressing a combination of T1R taste receptor sequences of the invention can also be used for receptor assays. Such expression can be used to determine whether a test compound specifically binds to a mammalian taste transmembrane receptor complex *in vivo* by contacting a non-human animal stably or transiently transfected with nucleic acids encoding chemosensory receptors or ligand-binding regions thereof with a test compound and determining whether the animal reacts to the test compound by specifically binding to the receptor polypeptide complex.

Animals transfected or infected with the vectors of the invention are particularly useful for assays to identify and characterize taste stimuli that can bind to a specific or sets of receptors. Such vector-infected animals expressing human taste receptor sequences can be used for *in vivo* screening of taste stimuli and their effect on, *e.g.*, cell physiology (*e.g.*, on taste neurons), on the CNS, or behavior. Alternatively, stable cell lines that express a T1R or combination thereof, can be used as nucleic transfer donors to produced cloned transgenic animals that stably express a particular T1R or combination. Methods of using nucleic transfer to produce cloned animals that express a desired heterologous DNA are the subject of several issued U.S. patents granted to the University of Massachusetts (licensed to Advanced Cell Technology, Inc.) and Roslin Institute (licensed to Geron Corp.).

Means to infect/express the nucleic acids and vectors, either individually or as libraries, are well known in the art. A variety of individual cell, organ, or whole animal parameters can be measured by a variety of means. The T1R sequences of the

invention can be for example co-expressed in animal taste tissues by delivery with an infecting agent, *e.g.*, adenovirus expression vector.

The endogenous taste receptor genes can remain functional and wild-type (native) activity can still be present. In other situations, where it is desirable that all taste receptor activity is by the introduced exogenous hybrid receptor, use of a knockout line is preferred. Methods for the constitutive of non-human transgenic animals, particularly transgenic mice, and the selection and preparation of recombinant constructs for generating transformed cells are well known in the art.

Constitutive of a "knockout" cell and animal is based on the premise that the level of expression of a particular gene in a mammalian cell can be decreased or completely abrogated by introducing into the genome a new DNA sequence that serves to interrupt some portion of the DNA sequence of the gene to be suppressed. Also, "gene trap insertion" can be used to disrupt a host gene, and mouse embryonic stem (ES) cells can be used to produce knockout transgenic animals (*see, e.g.*, Holzschu, *Transgenic Res* 6:97-106 (1997)). The insertion of the exogenous is typically by homologous recombination between complementary nucleic acid sequences. The exogenous sequence is some portion of the target gene to be modified, such as exonic, intronic or transcriptional regulatory sequences, or any genomic sequence which is able to affect the level of the target gene's expression; or a combination thereof. Gene targeting via homologous recombination in pluripotential embryonic stem cells allows one to modify precisely the genomic sequence of interest. Any technique can be used to create, screen for, propagate, a knockout animal, *e.g.*, *see* Bijvoet, *Hum. Mol. Genet.* 7:53-62 (1998); Moreadith, *J. Mol. Med.* 75:208-216 (1997); Tojo, *Cytotechnology* 19:161-165 (1995); Mudgett, *Methods Mol. Biol.* 48:167-184 (1995); Longo, *Transgenic Res.* 6:321-328 (1997); U.S. Patents Nos. 5,616,491; 5,464,764; 5,631,153; 5,487,992; 5,627,059; 5,272,071; WO 91/09955; WO93/09222; WO 96/29411; WO 95/31560; WO 91/12650.

The nucleic acids of the invention can also be used as reagents to produce "knockout" human cells and their progeny. Likewise, the nucleic acids of the invention can also be used as reagents to produce "knock-ins" in mice. The human or rat T1R gene sequences can replace the orthologous T1R in the mouse genome. In this way, a mouse expressing a human or rat T1R is produced. This mouse can then be used to analyze the function of human or rat T1Rs, and to identify ligands for such T1Rs.

a. Modulators

The compounds tested as modulators of a T1R family member can be any small chemical compound, or a biological entity, such as a protein, nucleic acid or lipid. Examples thereof include 5' IMP and 5' GMP. Essentially any chemical compound can be used as a potential modulator or ligand in the assays of the invention, although most often compounds that are soluble in aqueous solutions are tested. Assays can be designed to screen large chemical libraries by automating the assay steps and providing compounds from any convenient source; these assays are typically run in parallel (*e.g.*, in microtiter formats on microtiter plates in robotic assays). It will be appreciated that chemical libraries can be synthesized by one of many chemical reactions (e.g. Senomyx proprietary chemistries). Additionally, there are many suppliers of chemical compounds, including Sigma (St. Louis, MO), Aldrich (St. Louis, MO), Sigma-Aldrich (St. Louis, MO), Fluka Chemika-Biochemica Analytika (Buchs, Switzerland) and the like.

In one preferred embodiment, high throughput screening methods involve providing a combinatorial chemical or peptide library containing a large number of potential taste affecting compounds (potential modulator or ligand compounds). Such "combinatorial chemical libraries" or "ligand libraries" are then screened in one or more assays, as described herein, to identify those library members (particular chemical species or subclasses) that display a desired characteristic activity. The compounds thus identified can serve as conventional "lead compounds" or can themselves be used as potential or actual taste modulators.

Preferably, such libraries will be screened against cells or cell lines that stably express a T1R or combination of T1Rs, i.e. T1R1/T1R3 or T1R2/T1R3 and preferably a suitable G protein, e.g. $G_{\alpha 15}$. As shown in the examples infra, such stable cell lines exhibit very pronounced responses to taste stimuli, e.g. umami or sweet taste stimuli. However, cells and cell lines that transiently express one or more T1Rs may also be used in such assays.

A combinatorial chemical library is a collection of diverse chemical compounds generated by either chemical synthesis or biological synthesis, by combining a number of chemical "building blocks" such as reagents. For example, a linear combinatorial chemical library such as a polypeptide library is formed by combining a set of chemical building blocks (amino acids) in every possible way for a given compound length (*i.e.*,

the number of amino acids in a polypeptide compound). Thousands to millions of chemical compounds can be synthesized through such combinatorial mixing of chemical building blocks.

Preparation and screening of combinatorial chemical libraries is well known to those of skill in the art. Such combinatorial chemical libraries include, but are not limited to, peptide libraries (*see, e.g.*, U.S. Patent 5,010,175, Furka, *Int. J. Pept. Prot. Res.*, 37:487-493 (1991) and Houghton *et al.*, *Nature*, 354:84-88 (1991)). Other chemistries for generating chemically diverse libraries can also be used. Such chemistries include, but are not limited to: peptoids (*e.g.*, PCT Publication No. WO 91/19735), encoded peptides (*e.g.*, PCT Publication WO 93/20242), random bio-oligomers (*e.g.*, PCT Publication No. WO 92/00091), benzodiazepines (*e.g.*, U.S. Pat. No. 5,288,514), diversomers such as hydantoins, benzodiazepines and dipeptides (Hobbs *et al.*, *Proc. Nat. Acad. Sci.*, 90:6909-6913 (1993)), vinylogous polypeptides (Hagihara *et al.*, *J. Amer. Chem. Soc.*, 114:6568 (1992)), nonpeptidal peptidomimetics with glucose scaffolding (Hirschmann *et al.*, *J. Amer. Chem. Soc.*, 114:9217-9218 (1992)), analogous organic syntheses of small compound libraries (Chen *et al.*, *J. Amer. Chem. Soc.*, 116:2661 (1994)), oligocarbamates (Cho *et al.*, *Science*, 261:1303 (1993)), peptidyl phosphonates (Campbell *et al.*, *J. Org. Chem.*, 59:658 (1994)), nucleic acid libraries (Ausubel, Berger and Sambrook, all *supra*), peptide nucleic acid libraries (U.S. Patent 5,539,083), antibody libraries (Vaughn *et al.*, *Nature Biotechnology*, 14(3):309-314 (1996) and PCT/US96/10287), carbohydrate libraries (Liang *et al.*, *Science*, 274:1520-1522 (1996) and U.S. Patent 5,593,853), small organic molecule libraries (benzodiazepines, Baum, *C&EN*, Jan 18, page 33 (1993); thiazolidinones and metathiazanones, U.S. Patent 5,549,974; pyrrolidines, U.S. Patents 5,525,735 and 5,519,134; morpholino compounds, U.S. Patent 5,506,337; benzodiazepines, 5,288,514, and the like).

Devices for the preparation of combinatorial libraries are commercially available (*see, e.g.*, 357 MPS, 390 MPS (Advanced Chem Tech, Louisville KY), Symphony (Rainin, Woburn, MA), 433A (Applied Biosystems, Foster City, CA), 9050 Plus (Millipore, Bedford, MA)). In addition, numerous combinatorial libraries are themselves commercially available (*see, e.g.*, ComGenex, Princeton, NJ; Tripos, Inc., St. Louis, MO; 3D Pharmaceuticals, Exton, PA; Martek Biosciences; Columbia, MD; *etc.*).

In one aspect of the invention, the T1R modulators can be used in any food product, confectionery, pharmaceutical composition, or ingredient thereof to thereby modulate the taste of the product, composition, or ingredient in a desired manner. For instance, T1R modulators that enhance sweet taste sensation can be added to sweeten a product or composition; T1R modulators that enhance umami taste sensation can be added to foods to increase savory tastes. Alternatively, T1R antagonists can be used to block sweet and/or umami taste.

b. Kits

T1R genes and their homologs are useful tools for identifying chemosensory receptor cells, for forensics and paternity determinations, and for examining taste transduction. T1R family member-specific reagents that specifically hybridize to T1R nucleic acids, such as T1R probes and primers, and T1R specific reagents that specifically bind to a T1R polypeptide, *e.g.*, T1R antibodies are used to examine taste cell expression and taste transduction regulation.

Nucleic acid assays for the presence of DNA and RNA for a T1R family member in a sample include numerous techniques are known to those skilled in the art, such as southern analysis, northern analysis, dot blots, RNase protection, S1 analysis, amplification techniques such as PCR, and *in situ* hybridization. In *in situ* hybridization, for example, the target nucleic acid is liberated from its cellular surroundings in such as to be available for hybridization within the cell while preserving the cellular morphology for subsequent interpretation and analysis. The following articles provide an overview of the art of *in situ* hybridization: Singer *et al.*, *Biotechniques*, 4:230250 (1986); Haase *et al.*, *Methods in Virology*, vol. VII, pp. 189-226 (1984); and *Nucleic Acid Hybridization: A Practical Approach* (Names *et al.*, eds. 1987). In addition, a T1R polypeptide can be detected with the various immunoassay techniques described above. The test sample is typically compared to both a positive control (*e.g.*, a sample expressing a recombinant T1R polypeptide) and a negative control.

The present invention also provides for kits for screening for modulators of T1R family members. Such kits can be prepared from readily available materials and reagents. For example, such kits can comprise any one or more of the following materials: T1R nucleic acids or proteins, reaction tubes, and instructions for testing

T1R activity. Optionally, the kit contains a biologically active T1R receptor or cell line that stably or transiently expresses a biologically active T1R containing taste receptor. A wide variety of kits and components can be prepared according to the present invention, depending upon the intended user of the kit and the particular needs of the user.

EXAMPLES

While the invention has been described in detail supra, the following examples are provided to illustrate preferred embodiments. These examples are intended to be illustrative and not limitative of the scope of the invention.

In the protein sequences presented herein, the one-letter code X or Xaa refers to any of the twenty common amino acid residues. In the DNA sequences presented herein, the one letter codes N or n refers to any of the of the four common nucleotide bases, A, T, C, or G.

Example 1

Production of Intronless hT1R Expression Constructs

Intronless hT1R expression constructs were cloned by a combination of cDNA-based and genomic DNA-based methods. To generate the full-length hT1R1 expression construct, two 5' coding exons identified in a cloned hT1R1 interval (accession # AL159177) were combined by PCR-overlap, and then joined to a 5'-truncated testis cDNA clone. The hT1R2 expression construct was generated from a partially sequenced hT1R2 genomic interval. Two missing hT1R2 5' exons were identified by screening shotgun libraries of the cloned genomic interval using probes derived from the corresponding rat coding sequence. Coding exons were then combined by PCR-overlap to produce the full-length expression construct. The hT1R3 expression construct was generated by PCR-overlap from a sequenced hT1R3 genomic interval (accession # AL139287). Rat T1R3 was isolated from a rat taste tissue-derived cDNA library using an rT1R3 exon fragment generated by hT1R3-based degenerate PCR. The partial hT1R1 cDNA, rT1R2 cDNA, and partial hT1R2 genomic sequences were obtained from Dr. Charles Zuker (University of California, San Diego).

The nucleic acid and amino acid sequences for the above-identified T1R cloned sequences as well as other full-length and partial T1R sequences are set forth in the sequence listing.

Also, the following conceptual translations, which correspond to the C-termini of two orthologous pairs of fish T1Rs, are derived from unpublished genomic sequence fragments and provided. Fugu T1RA was derived from accession 'scaffold 164'; Fugu T1RB was derived from accession LPC61711; Tetradon T1RA was derived from accession AL226735; Tetradon T1RB was derived from accession AL222381. Ambiguities in the conceptual translations ('X') result from ambiguities in database sequences. These sequences can be found in the sequence listing.

Additionally, the accession number and reference citations relating to mouse and rat T1Rs and allelic variants thereof in the public domain are is set forth below: rT1R1 (Accession # AAD18069) (Hoon et al., Cell 96 (4): 541-51 (1999)); rT1R2 (Accession # AAD18070) (Hoon et al., Cell 96(4): 541-59 (1999)); mT1R1 (Accession # AAK39437); mT1R2 (Accession #AAK 39438); mT1R3 (Accession AAK 55537) (Max et al., Nat. Genet. 28(l): 58-63 (2001)); rT1R1 (Accession # AAK7092) (Li et al., Mamm. Genome (12(l): 13-16 (2001)); mT1R1 (Accession # NP 114073); mT1R1 (Accession # AAK07091) (Li et al., Mamm. Genome (121):13-16 (2001)); rT1R2 (Accession # AAD18070) (Hoon et al., Cell 9664): 541-551 (1999)); mT1R2 (Accession # NP114079); mT1R3 (Accession # AAK39436); mT1R3 (Accession # BAB47181); (Kitagawa et al., Biochem. Biophys. Res. Comm. 283(l):236-42 (2001)); mT1R3 (Accession #NP114078); mT1R3 (Accession # AAK55536) (Max et al., Nat. Genet. 28(l):58-63 (2001)); and mT1R3 (Accession No. AAK01937).

Example 2

Sequence Alignment of Human and Rat T1Rs

Cloned T1R sequences selected from those identified above were aligned against the corresponding rat T1Rs. As shown in Figure 1, human T1R1, human T1R2 and human T1R3 and rat T1R3 were aligned with previously described T1Rs (rT1R1 having Accession # AAD18069 and rT1R2 having Accession # AAD18070), the rat mGluR1 metabotropic, glutamate receptor (Accession # P23385); and the human calcium-sensing receptor (Accession #P41180). For clarity of the comparison, the

mGluR1 and calcium-sensing receptor C-termini are truncated. The seven potential transmembrane segments are boxed in blue. Residues that contact the glutamate side-chain carbutylate in the mGluR1 crystal structure are boxed in red, and residues that contact the glutamate α-amino acid moiety are boxed in green. The mGluR1 and calcium-sensing receptor cysteine residues implicated in intersubunit disulfide-based formation are circled in purple. These cysteines are not conserved in T1R1 and T1R2, but are located in a degraded region of the alignment that contains a potentially analogous T1R3 cysteine residue, also circled.

Example 3

Demonstration by RT-PCR that hT1R2 and hT1R3 are expressed in taste tissue

As shown in Figure 2, hT1R2 and hT1R3 are expressed in taste tissue: expression of both genes can be detected by RT-PCR from resected human circumvallate papillae.

Example 4

Methods for Heterologous Expression of T1Rs in Heterologous Cells

An HEK-293 derivative (Chandrashekar et al., Cell 100(6): 703-11 (2000)), which stably expresses Gα15, was grown and maintained at 37°C in Dulbecco's Modified Eagle Medium (DMEM, Gibco BRL) supplemented with 10% FBS, MEM non-essential amino acids (Gibco BRL), and 3 μg/ml blasticidin. For calcium-imaging experiments, cells were first seeded onto 24-well tissue-culture plates (approximately 0.1 million cells per well), and transfected by lipofection with Mirus TransIt-293 (PanVera). To minimize glutamate-induced and glucose-induced desensitization, supplemented DMEM was replaced with low-glucose DMEM/GlutaMAX (Gibco BRL) approximately 24 hours after transfection. 24 hours later, cells were loaded with the calcium dye Fluo-4 (Molecular Probes), 3μM in Dulbecco's PBS buffer (DPBS, GibcoBRL), for 1.5 hours at room temperature. After replacement with 250μl DPBS, stimulation was performed at room temperature by addition of 200μl DPBS supplemented with taste stimuli. Calcium mobilization was monitored on a Axiovert S100 TV microscope (Zeiss) using Imaging Workbench 4.0 software (Axon). T1R1/T1R3 and T1R2/T1R3 responses were strikingly transient – calcium increases rarely persisted longer than 15 seconds – and asynchronous. The number of responding

cells was thus relatively constant over time; therefore, cell responses were quantitated by manually counting the number of responding cells at a fixed time point, typically 30 seconds after stimulus addition.

Example 5

Human T1R2/T1R3 functions as a sweet taste receptor

HEK cells stably expressing Gα15 were transiently transfected with human T1R2, T1R3 and T1R2/T1R3, and assayed for increases in intracellular calcium in response to increasing concentrations of sucrose (Figure 3(a)). Also, T1R2/T1R3 dose responses were determined for several sweet taste stimuli (Figure 3(b)). The maximal percentage of responding cells was different for different sweeteners, ranging from 10-30%. For clarity, dose responses were normalized to the maximal percentage of responding cells. The values in Figure 3 represent the mean ± s.e. of four independent responses. X-axis circles mark psychophysical detection thresholds determined by taste testing. Gurmarin (50-fold dilution of a filtered 10g/l *Gymnema sylvestre* aqueous extract) inhibited the response of T1R2/T1R3 to 250 mM sucrose, but not the response of endogenous β2-adrenergic receptor to 20 μM isoproterenol (Figure 3(b)). Figure 3(c) contains the normalized response of T1R2/T1R3 co-expressing cell lines to different sweeteners(sucrose, aspartame, D-tryptophan and saccharin).

Example 6

Rat T1R2/T1R3 also functions as a sweet taste receptor

HEK cells stably expressing Gα15 were transiently transfected with hT1R2/hT1R3, rT1R2/rT1R3, hT1R2/rT1R3, and rT1R2/hT1R3. These transfected cells were then assayed for increased intracellular calcium in response to 350 mM sucrose, 25 mM tryptophan, 15 mM aspartame, and 0.05% of monellin. The results with sucrose and aspartame are contained in Figure 4 and indicate that rT1R2/rT1R3 also functions as a sweet taste receptor. Also, these results suggest that T1R2 may control T1R2/T1R3 ligand specificity.

Example 7

T1R2/T1R3 responses using an automated fluorescence based assay

HEK cells stably expressing Gα15 were transiently transfected with hT1R2 and hT1R3. These cells were loaded with the calcium dye Fluo-4, and their responses to a sweetener measured using a fluorescence plate reader. Figure 5 contains cyclamate (12.5 mM) responses for cells expressing hT1R2/hT1R3 and for cells expressing only hT1R3 (J19-22). The fluorescence results obtained indicate that responses to these taste stimuli only occurred in the cells expressing hT1R2/hT1R3. Figure 6 contains normalized dose-response curves, the results of which show that hT1R2 and hT1R3 function together as a human taste receptor based on their dose-specific interaction with various sweet stimuli . Particularly, Figure 6 contains dose-responses for sucrose, tryptophan and various other commercially available sweeteners. These results indicate that T1R2/T1R3 is a human sweet taste receptor as the rank order and threshold values obtained in the assay closely mirror values for human sweet taste.

Example 8

Ligand-binding residues of mGluR1 are conserved in T1R1

As shown in Figure 6, the key ligand-binding residues of mGluR1 are conserved in T1R1. The interaction of glutamate with mGluR1 is shown with several key residues highlighted according to the same color scheme as Figure 1.

Example 9

Human T1R1/T1R3 functions as umami taste receptors

HEK cells stably expressing Gα15 were transiently transfected with human T1R1, T1R3 and T1R1/T1R3 and assayed for increases in intracellular calcium in response to increasing concentrations of glutamate (Figure 8(a)), and 0.5 mM glutamate), 0.2 mM IMP, and 0.5 mM glutamate plus 0.2 mM IMP (Figure 8(b)). Human T1R1/T1R3 dose responses were determined for glutamate in the presence and absence of 0.2 mM IMP (Figure 8(c)). The maximal percentages of responding cells was approximately 5% for glutamate and approximately 10% for glutamate plus IMP. For clarity, does responses are normalized to the maximal percentage of responding cells. The values represent the mean ± s.e. of four independent responses. X-axis circles mark taste detection thresholds determined by taste testing.

Example 10

PDZIP as an Export Sequence

The six residue PDZIP sequence (SVSTW (SEQ ID NO:1)) was fused to the C-terminus of hT1R2 and the chimeric receptor (i.e. hT1R2-PDZIP) was transfected into an HEK-293 host cell. The surface expression of hT1R2 was then monitored using immunofluorescence and FACS scanning data. As shown in Figures 9A and 9B, the inclusion of the PDZIP sequence increased the surface expression of hT1R2-PDZIP relative to hT1R2. More specifically, Figure 9A shows an immunofluorescence staining of myc-tagged hT1R2 demonstrating that PDZIP significantly increases the amount of hT1R2 protein on the plasma membrane. Figure 9B shows FACS analysis data demonstrating the same result– Cells expressing myc-tagged hT1R2 are indicated by the dotted line and cells expressing myc-tagged hT1R2-PDZIP are indicated by the solid line. Particularly, Figure 10A shows untransfected Gα15 stable host cells in HBS

buffer, Figure 10B shows hT1R2-PDZIP transfected Gα15 stable hose cells in sweetener pool no. 5 (saccharin, sodium cyclamate, Acesulfame K, and Aspartame-20 mM each in HBS buffer), Figure 10C shows T1R3-PDZIP transfected Gα15 stable host cells in sweetener pool no. 5, and Figure 10D shows hT1R2-PDZIP/hT1R3-PDZIP co-transfected Gα15 stable host cells in sweetener pool no. 5. Further, Figures 10E-10H show dose-dependent response of hT1R2/hT1R3 co-transfected Gα15 stable host cells to sucrose-E: 0mM in HBS buffer; F: 30 mM; G: 60 mM; and H: 250 mM. Figures 10I-10L shown the responses of hT1R2/hT1R3 co-transfected Gα15 stable host cells to individual sweeteners – I: Aspartame (1.5 mM); J: Acesulfame K (1 mM); K: Neotame (20mM); L: Sodium cyclamate (20mM). As demonstrated by the calcium-images of Figure 10, hT1R2 and hT1R3 are both required for the activities triggered by the sweet stimuli.

Example 11

Generation of Cell Lines that Stably Co-Express T1R1/T1R3 or T1R2/T1R3

Human cell lines that stably co-express human T1R2/T1R3 or human T1R1/T1R3 were generated by transfecting linearized PEAK10-derived (Edge Biosystems) vectors and pCDNA 3.1/ZEO-derived (Invitrogen) vectors respectively containing hT1R1 or hT1R2 expression construct (plasmid SAV2485 for T1R1, SAV2486 for T1R2) and hT1R3 (plasmid SXV550 for T1R3) into a $G_{\alpha 15}$ expressing cell line. Specifically, T1R2/T1R3 stable cell lines were produced by co-transfecting linearized SAV2486 and SXV550 into Aurora Bioscience's HEK-293 cell line that stably expresses $G_{\alpha 15}$. T1R1/T1R3 stable cell lines were produced by co-transfecting linearized SAV2485 and SXV550 into the same HEK-293 cell line that stably expresses $G_{\alpha 15}$. Following SAV2485/SXV550 and SAV2486/SXV550 transfections, puromycin-resistant and zeocin-resistant colonies were selected, expanded, and tested by calcium imaging for responses to sweet or umami taste stimuli. Cells were selected in 0.0005 mg/ml puromycin (CALBIOCHEM) and 0.1 mg/ml zeocin (Invitrogen) at 37°C in low-glucose DMEM supplemented with GlutaMAX, 10% dialyzed FBS, and 0.003 mg/ml blasticidin. Resistant colonies were expanded, and their responses to sweet taste stimuli evaluated by Fluorescence microscopy. For automated fluorimetric imaging on VIPR-II instrumentation (Aurora Biosciences), T1R2/T1R3 stable cells were first seeded onto 96-well plates (approximately 100,000 cells per well). Twenty-

four hours later, cells were loaded with the calcium dye fluo-3-AM (Molecular Probes), 0.005 mM in PBS, for one hour at room temperature. After replacement with 70 μl PBS, stimulation was performed at room temperature by addition of 70 μl PBS supplemented with taste stimuli. Fluorescence (480 nm excitation and 535 nm emission) responses from 20 to 30 seconds following compound addition were averaged, corrected for background fluorescence measured prior to compound addition, and normalized to the response to 0.001 mM ionomycin (CALBIOCHEM), a calcium ionophore.

It was then observed that when these cell lines were exposed to sweet or umami stimuli, that for active clones typically 80-100% of cells responded to taste stimuli. Unexpectedly, the magnitude of individual cell responses was markedly larger than that of transiently transfected cells.

Based on this observation, the inventors tested the activity of T1R stable cell lines by automated fluorescence imaging using Aurora Bioscience's VIPR instrumentation as described above. The responses of two T1R1/T1R3 and one T1R2/T1R3 cell line are shown in Figure 11 and Figure 12 respectively.

Remarkably, the combination of increased numbers of responding cells and increased response magnitudes resulted in a greater than 10-fold increase in activity relative to transiently transfected cells. (By way of comparison, the percent ionomycin response for cells transiently transfected with T1R2/T1R3 was approximately 5% under optimal conditions.) Moreover, dose responses obtained for stably expressed human T1R2/T1R3 and T1R1/T1R3 correlated with human taste detection thresholds. The robust T1R activity of these stable cell lines suggests that they are well suited for use in high-throughput screening of chemical libraries in order to identify compounds, e.g. small molecules, that modulate the sweet or umami taste receptor and which therefore modulate, enhance, block or mimic sweet or umami taste.

Example 12

Generation of cell lines that inducibly co-express T1R1/T1R3 which selectively respond to umami taste stimuli

T1R1/T1R3 HEK 293 cell lines that stably expressed the umami taste receptor display robust improved activity relative to transiently tranfected cites. However, a disadvantage is that they can rapidly lose activity during cell propagation.

Also, these findings show that (i) T1R1/T1R3 is a umami taste receptor, i.e., and (ii) that cell lines which robustly express T1R1/T1R3, preferably stable and/or inducible T1R1/T1R3 cell lines can be used in assays, preferably for high throughput screening of chemical libraries to identify novel modulators of umami taste. Modulators that enhance umami taste may be used.

To overcome the instability of the T1R1/T1R3 stable cell lines, the HEK-$G_{\alpha 15}$ cells have been engineered to inducibly express T1R1/T1R3 using the GeneSwitch system (Invitrogen). pGene-derived zeocin-resistant expression vectors for human T1R1 and T1R3 (plasmid SXV603 for T1R1 and SXV611 for T1R3) and a puromycin-resistant pSwitch-derived vector that carries the GeneSwitch protein (plasmid SXV628) were linearized and cotransfected into the HEK-$G_{\alpha 15}$ cell line. Zeocin-resistant and puromycin-resistant colonies were selected, expanded, induced with variable amounts of mifepristone, and tested by calcium imaging for responses to umami taste stimuli.

Inducible expression of T1R1/T1R3 resulted in robust activity. For example, approximately 80% of induced cells but only approximately 10% of transiently transfected cells responded to L-glutamate; More specifically, pGene derived Zeocin-resistant expression vectors that express human T1R1 and human T1R3 and a puromycn-resistant pSwitch-derived vector that carries the GeneSwitch protein were linearized and co-tranfected into $G_{\alpha 15}$ cells. Cells were selected in 0.5 µg/ml puromycin (CAL BIOCHEM) and 100 µg/ml Zeocin (Invitrogen) at 37°C in Dulbecco's Modified Eagle Medium supplemented with GlutaMAX, (10 % dialyzed FBS, and 3 ug/ml blasticidin. Resistant colonies were expanded, and their responses to umami taste stimuli following induction with 10^{-10} M mifepristone determined by fluorescence microscopy following the methods of Li et al., PNAS 99(7): 4692-4696 (2002).

For automated fluorometric imaging on FLIPR instrumentation (Molecular Device), cells from one clone (designated clone I-17) were seeded into 96-well plates (approximately 80,000 cell per well) in the presence of 10^{-10} M mifepristone and incubated for 48 hours. Cells were then loaded with the calcium dye fluo-4-AM (Molecular Probes), 3 µM in PBS, for 1.5 hours at room temperature.

After replacement with 50 µl PBS, stimulation was performed at room temperature by the addition of 50 µl PBS supplemented with different stimuli. In contrast to previous transient T1R1/T1R3 umami receptor expression systems that

necessitated quantifying T1R1/T1R3 receptor activity by individually counting responding cells (Li et al., PNAS 99(7): 4692-4696 (2002)) (because of the low activity of the receptor therein), the subject inducible expression system resulted in a clone I-17 having substantially increased activity that allowed receptor activity to be quantified by determining maximal fluorescence increases (480 nm excitation and 535 nm emission) summated over fields of imaged cells. The maximal fluorescence from four independent determinations was averaged, corrected for background fluorescence measured prior to compound addition, and normalized to the response to 0.002 mM ionomycin (CALBIOCHEM).

These results are contained in Figure 13. Particularly, Figure 13 contains a dose-response curve determined for L-glutamate in the presence and absence of 0.2 mM IMP. In the figure, each value represents average summated maximal fluorescence (corrected for background fluorescence) for four independent determinations. These dose-response curves correspond to those determined for cells transiently transfected with T1R1/T1R3.

The selectivity of the umami T1R1/T1R3 taste receptor was also evaluated by screening with different L-amino acids. The results obtained indicated that T1R1/T1R3 is selectively activated by the umami-tasting L-amino acids (L-glutamate and L-aspartate).

The results of experiments wherein the responses of the I-17 clone was resulted in tested in the presence of different L-amino acids are contained in Figure 14 and Figure 15. Figure 14 shows the results of an experiment wherein the I-17 cell line was contacted with different L-amino acids at a concentration of 10mM in the presence and absence of 1mM IMP.

Figure 15 contains a dose-response curve for active amino acids determined in the presence of 0.2mM IMP. Each value represents the average of four independent determinations.

The results obtained in these experiments support the specificity and selectivity of the umami taste receptor to umami taste stimuli. Whereas the umami taste stimuli L-glutamate and L-aspartate significantly activated the T1R1/T1R3 receptor at different concentrations (see Figure 14 and 15), the other L-amino acids which activated the human T1R1/T1R3 receptor only activated the receptor weakly and at much higher concentrations.

Therefore, these results support the selectivity of the T1R1/T1R3 receptor for umami taste stimuli and the suitability of this inducible stable expression system for use in high throughput screening assays using automated fluorometric imaging instrumentation to identify compounds that activate the umami taste receptor, for example L-glutamate or L-aspartate , or which enhance the activity of L-glutamate to activate the umami taste receptor, for example 5'-IMP or 5'-GMP, or block the activation of the umami taste receptor by umami taste stimuli such as L-glutamate and L-aspartate.

Compounds identified using these assays have potential application as flavorants in foods and beverage compositions for mimicing or blocking umami taste stimuli.

Example 13

Lactisole Inhibits the Receptor Activities of Human T1R2/T1R3 and T1R1/T1R3 , and Sweet and Umami Taste

Lactisole, an aralkyl carboxylic acid, was thought to be a selective sweet-taste inhibitor (See e.g., Lindley (1986) U.S. Patent 4,567,053; and Schiffman et al. Chem Senses 24:439-447 (1999)). Responses of HEK-$G_{\alpha15}$ cells transiently transfected with T1R2/T1R3 to 150 mM sucrose in the presence of variable concentrations of lactisole were measured. Lactisole inhibits the activity of human T1R2/T1R3 with an IC_{50} of 24 μM.

The T1R1/T1R3 umami and T1R2/T1R3 sweet taste receptor may share a common subunit. It has therefore been theorized that lactisole, which inhibit the T1R2/T1R3 sweet taste receptor, may have a similar effect on the T1R1/T1R3 umami taste receptor. The present inventors tested the effect of lactisole on the response of human T1R1/T1R3 to 10mM L-Glutamate. As with the T1R2/T1R3 sweet receptor, lactisole inhibited T1R1/T1R3 with an IC_{50} of 165 μM. Lactisole inhibition likely reflects antagonism at the T1R receptors instead of, for example, non-specific inhibition of $G_{\alpha15}$-mediated signaling because the response of muscarinic acetylcholine receptors was not inhibited by lactisole.

The present inventors then evaluated the effect of lactisole on human umami taste. Taste thresholds in the presence of 1 and 2 mM lactisole were determined for the umami taste stimuli L-Glutamate with or without 0.2 mM IMP, the sweet taste stimuli

sucrose and D-tryptophan, and the salty taste stimulus sodium chloride following the methods of Schiffman et al. (Chem. Senses 24: 439-447 (1989)). Millimolar concentrations of lactisole dramatically increased detection thresholds for sweet and umami but not salt taste stimuli. These results are contained in Figure 16.

In conclusion, (i) these findings further support the inventors' hypothesis that T1R1/T1R3 is the only umami taste receptor, and (ii) the T1R1/T1R3 and T1R2/T1R3 receptors may share a structurally related lactisole-binding domain.

While the foregoing detailed description has described several embodiments of the present invention, it is to be understood that the above description is illustrative only and not limiting of the disclosed invention. The invention is to be limited only by the claims which follow.

Example 14

Mapping of Ligand Interaction Sites on the Sweet Receptor

Through coexpression of T1R2R-H with human T1R3, part of the human sweet receptor (the N-terminal domain of T1R2) was replaced with rat protein sequence. The responses to aspartame and neotame are abolished, showing that the N- terminal domain of human T1R2 is required for recognizing aspartame and neotame. Similarly, the rat T1R2 N-terminal domain was also replaced with human protein sequence by coexpressing T1R2H-R with rat T1R3. The chimeric receptor gains the ability to respond to aspartame and neotame, suggesting that the same domain of human T1R2 is also sufficient (in the context of sweet receptors) to recognize those two sweeteners (Fig. 22B). These *in vitro* functional expression data indicate that the important interaction determinants are located in the N-terminal extracellular domain.

In contrast, replacing either half of human T1R2 with rat protein sequence does not affect its response to cyclamate. Instead, the C-terminal domain of human T1R3 is required and sufficient, when co-expressed with T1R2, to recognize cyclamate (Fig. 22C). The transmembrane domain of family C GPCRs has been known to contain binding sites for allosteric modulators (Gasparini, F., R. Kuhn, and J.P. Pin, Curr Opin Pharmacol 2002 Feb;2(1):43-9). This is the first case in family C GPCR, where an agonist binds directly to the transmembrane domain and activates the receptor in the absence of other ligand. .

Lactisole, an aralkyl carboyxlic acid, is a specific human sweet taste inhibitor, which has physiological effect on the rodent taste. Consistent with the taste effect, lactisole inhibits the human but not rat T1R2/T1R3 response to sucrose in our assay system (Fig. 22A). The same kind of mapping experiments on lactisol interaction site using the T1R chimeras was performed. Like cyclamate, lactisole requires the human T1R3 C-terminal domain to inhibit the receptor's response to sucrose and acesulfame K

(Fig. 22D). This result further demonstrates the importance of T1R3 C-terminal domain in the sweet receptor function. The chimeras in all 16 possible combinations were tested, and all functional combinations generated results consistent with our model.

Mutagenesis studies were conducted on both T1R2 and T1R3 to narrow down the essential amino acids in recognition of aspartame, neotame, and cyclamate. If T1R2 and T1R3 are responsible for recognizing different sweeteners, mutations in T1R2 N-terminal domain would affect responses to aspartame and neotame, but not cyclamate. In addition, mutations in T1R3 C-terminal domain would have the opposite effect. To select the crucial amino acid residues in the T1R2 N-terminal domain, the sequence of T1R2 was aligned with mGluR1 (Fig. 23A). Among the eight residues that are crucial in ligand binding in mGluR1 (Kunishima, N., et al., Nature, 2000. 407(6807): p. 971-7), three are conserved in human T1R2 (S144, Y218, and E302). Each of the three residues were mutated and the resulting receptors were tested for their response to different sweeteners. Substitution of Y218 to A abolished the responses to all sweeteners tested, showing Y218 is important for the overall conformation of the receptor. The two other hT1R2 variants, containing S144A and E302A, selectively affected the response to aspartame and neotame but not cyclamate. Stable cell lines expressing S144A and E302A hT1R2 variants (coexpressed with wild type hT1R3 and $G_{\alpha 15}$) did not respond to aspartame or neotame at the physiological concentrations, but did respond to cyclamate (Fig. 23B).

In order to further map the cyclamate-binding site, the three extracellular loops in the T1R3 C-terminal domain were focused on. Alignment of human and rodent T1R3s reveal multiple amino acid differences in the three extracellular loops (Fig. 23C). Replacing extracellular loop-2 or loop-3 with rat sequences abolished the cyclamate response without affecting the sucrose or aspartame responses. In contrast,

replacing extracellular loop 1 had no obvious effect on response to cyclamate, showing an important role for EC loops 2 and 3 in recognizing cyclamate (Fig. 23D). None of those loop-replacements affected the inhibition effect of lactisole, showing a different binding mechanism. In summary, amino acid substitutions in T1R2 or T1R3 result in selective interference of activities induced by different sweeteners, consistent with the chimeric receptor results.

The above results demonstrate that the human sweet receptor function as a heteromeric complex of T1R2 and T1R3. Both subunits are required for recognizing different sweeteners, and the data indicate the existence of multiple binding pockets on the receptor for different classes of agonists. The presence of multiple ligand-binding sites provides structural guidance and definition for the specifically binding compounds of the invention.

Example 15

Mapping of Receptor-G protein Interactions

The human and rat sweet receptors are also different in their G protein-coupling efficiency. Even though both human and rat receptors can couple efficiently to $G_{\alpha 15/i1}$, only the human receptor can couple efficiently to $G_{\alpha 15}$ (Fig. 24A). This species difference allows for mapping of the receptor G protein interactions using the same chimeric receptors as described above. T1R2 but not T1R3 appears to be critical for $G_{\alpha 15}$-coupling, since replacing the C-terminus of human T1R2 with the corresponding rat sequence abolished coupling, and replacing rat T1R2 C-terminal half with human sequence enabled the receptor to couple to $G_{\alpha 15}$ and respond to sucrose and acesulfame K (Fig. 24); Swapping the T1R3 C-terminal sequences had no effect on $G_{\alpha 15}$–coupling (Fig. 24B). This observation demonstrates the important role of T1R2 in G protein-coupling in the functional expression system. Gustducin (Wong, G.T., K.S. Gannon, and R.F. Margolskee, Nature, 1996. 381(6585): p. 796-800) has been proposed to be an endogenous G protein for the sweet taste receptor, and T1R2 can be the subunit responsible for *in vivo* coupling in taste cells. GABA$_B$R is the other example of heteromeric family C GPCR, whereas one subunit (GABA$_B$R1) is responsible for ligand- binding, and the other (GABA$_B$R2) for G protein coupling (Margeta-Mitrovic, M., Paroc Natl Acad Sci U S A, 2001. 98(25): p. 14643-8; Margeta-Mitrovic, M., Proc Natl Acad Sci U S A, 2001. 98(25): p. 14649-54). The sweet receptor is different from GABA$_B$R in that T1R2 is required for both ligand recognition and G- protein coupling.

Example 16

Lactisole Antagonizes Human T1R1/T1R3 and Inhibits Human Umami Taste

It was hypothesized that since T1R1/T1R3 function as heteromeric receptors as well as the sweet receptor, that lactisole should have similar effect on T1R1/T1R3 activity, since T1R3 is a common subunit between the sweet and the umami receptors. Indeed, lactisole antagonized human T1R1/T1R3 (Fig. 25A). Lactisole acts as a noncompetitive inhibitor of T1R1/T1R3, since the IC_{50} values are apparently not dependent on glutamate concentration (Fig. 25B), and lactisole reduces the maximal activities of the receptor without significantly changing the EC_{50} of agonists (Fig. 25C). These results demonstrate that lactisole binds to a different site from L-glutamate, and are consistent with the hypothesis that the glutamate-binding pocket is located in T1R1. Lactisole appears to be a competitive inhibitor of the sweet receptor, as its IC_{50}s are dependent on the concentrations of the sweeteners, and it increases the EC_{50}s of the sweeteners without significantly affecting the maximal activities.

The inhibition effect of lactisole is mediated by the T1R receptors since it had no effect on the endogenous muscarinic acetylcholine receptor in HEK cells or on a mouse bitter receptor, mT2R5, transiently expressed in HEK cells. As was the case for the T1R2/T1R3 receptor, lactisole inhibition of the T1R1/T1R3 response to umami taste stimuli was reversible following washout and restimulation.

To correlate the receptor activity with behavior, the effect of lactisole on human umami taste was tested. As predicted, millimolar concentrations of lactisole dramatically increased detection thresholds for sweet and umami but not salt taste stimuli (Fig. 25D). Lactisole was previously not known to be an umami taste inhibitor. The correlation between receptor activity and taste results demonstrates a crucial role of T1Rs in human umami taste.

Example 17

Cyclamate Enhances Human T1R1/T1R3 Receptor Activities

Based on the same heteromeric model of T1Rs (Fig. 26), it was predicted that cyclamate would also modulate the activity of the human T1R1/T1R3 umami receptor by acting on T1R3. Although cyclamate alone had no effect on T1R1/T1R3, it enhanced the activity of the receptor in the presence of L-glutamate (Fig. 27E). This

effect is specific for the human T1R1/T1R3, as cyclamate had no effect on the activities of the endogenous muscarinic acetylcholine receptor in the presence of carbachol (Fig. 27E). It is noteworthy that cyclamate has comparable EC_{50}s for the sweet receptor (Fig. 23B) and umami receptor. Cyclamate reproducibly left-shifts the dose-response curves for L-glutamate by ~2 fold in the presence or absence of IMP (Fig. 25F). IMP has a more dramatic effect of enhancing the receptor, and the effect of cyclamate is observed in the presence of IMP (Fig. 25F), suggesting a different mechanism from IMP in enhancing the receptor. IMP appears to bind to T1R1, since it has no effect on the sweet receptor. Other sweeteners, including sucrose, aspartame, saccharin, and D-tryptophan, had no effect on the human T1R1/T1R3 activities.

In summary, it has been demonstrated that both T1R2 and T1R3 are required in a functional sweet receptor, that aspartame and neotame require the N-terminal extracellular domain of T1R2, G protein-coupling requires C-terminal half of T1R2, and that cyclamate and lactisole require the transmembrane domain of T1R3. These findings demonstrate the different functional roles of T1R subunits in a heteromeric complex and the presence of multiple sweetener interaction sites on the sweet receptor. Because T1R3 is the common subunit in the sweet and the umami receptors, it was predicted and confirmed the effect of cyclamate and lactisole on the umami receptor. Furthermore, a correlation was able to be made between the lactisole effect on the receptor activities with taste. Based on these observations, a model was created (Fig. 26) for the structure-function relationships of the T1R family taste receptors. Natural carbohydrate sweeteners bind to the N-terminal domain of T1R2, similar to aspartame and neotame, and there are other ligand binding sites on the sweet receptor as well, for example, the transmembrane domain of T1R2. The umami receptor functions similarly as a heteromeric complex, and MSG and IMP each appears to bind to the T1R1 subunit, since neither has any effect on the sweet receptor, and the transmembrane domain of T1R1 is responsible for coupling to G proteins.

Example 18

HTS Protocol for Sweet Tastants

An HEK293 cell line derivative (Chandrashekar, J., Mueller, K.L., Hoon, M.A., Adler, E., Feng, L., Guo, W., Zuker, C.S., Ryba, N.J.,. *Cel,l* 2000, *100*, 703-711.) that stably expresses Gα15 and hT1R2/hT1R3 (Li, X., Staszewski, L., Xu, H., Durick, K., Zoller, M., Adler, E. *Proc Natl Acad Sci U S A* 2002, *99*, 4692-4696, World Patent # WO 03/001876 A2, herein incorporated by reference in their entirety) was used to identify compounds with sweet taste enhancing properties.

Compounds were initially selected based on their activity on the hT1R2/hT1R3-HEK293-Gα15 cell line (Li, et al. *vide supra*). Activity was determined using an automated fluorometric imaging assay on a FLIPR instrument (Fluorometric Intensity Plate Reader, Molecular Devices, Sunnyvale, CA) (designated FLIPR assay). Cells from one clone (designated S-9 cells) were seeded into 384-well plates (at approximately 50,000 cells per well) in a medium containing DMEM Low Glucose (Invitrogen, Carlsbad, CA), 10% dialyzed fetal bovine serum (Invitrogen, Carlsbad, CA), 100 Units/ml Penicillin G, and 100 μg/ml Streptomycin (Invitrogen, Carlsbad, CA) (Li, et al. *vide supra*) see also World Patent #WO 03/001876 A2). S-9 cells were grown for 24 hours at 37 °C. S-9 cells were then loaded with the calcium dye Fluo-3AM (Molecular Probes, Eugene, OR), 4 μM in a phosphate buffered saline (D-PBS) (Invitrogen, Carlsbad, CA), for 1 hour at room temperature. After replacement with 25 μl D-PBS, stimulation was performed in the FLIPR instrument and at room temperature by the addition of 25 μl D-PBS supplemented with different stimuli at concentrations corresponding to twice the desired final level. Receptor activity was quantified by determining the maximal fluorescence increases (using a 480 nm excitation and 535 nm emission) after normalization to basal fluorescence intensity measured before stimulation.

For dose-responses analysis, stimuli were presented in duplicates at 10 different concentrations ranging from 60 nM to 30 μM. Activities were normalized to the response obtained with 400 mM D-fructose, a concentration that elicits maximum receptor response. EC_{50}s were determined using a non-linear regression algorithm

(using Senomyx, Inc. software), where the Hill slope, bottom asymptotes and top asymptotes were allow to vary. Identical results were obtained when analyzing the dose-response data using commercially available software for non-linear regression analysis such as GraphPad PRISM (San Diego, CA).

In order to determine the dependency of hT1R2/hT1R3 for the cell response to different stimuli, selected compounds were subjected to a similar analysis on HEK293-Gα15 cells (not expressing the human sweet receptor). The HEK293-Gα15 cells do not show any functional response in the FLIPR assay to D-Fructose or any other known sweeteners. Similarly, compounds described herein do not induce any functional response when using HEK293-Gα15 cells in the FLIPR assay.

Example 19

Flavor Enhancement Measurements for Sweet Tastants using Human Volunteers

Basic screening of sensory taste testers: Potential panelists were tested for their abilities to rank and rate intensities of solutions representing the five basic tastes. Panelists ranked and rated intensity of five different concentrations of each of the five following compounds: sucrose (sweet), sodium chloride (salty), citric acid (sour), caffeine (bitter), and monosodium glutamate (umami). Panelists tasted a total of 25 samples per session (5 samples of each of the 5 solution types). In the first session, panelists ranked the five concentrations for intensity of the attribute in question. This was repeated four more times with other samples. In the second session, panelists rated intensity of the five concentrations of each sample using a line scale called the "Labeled Magnitude Scale" (LMS). The LMS is anchored with intensities (e.g. barely detectable, weak, moderate, strong, very strong, and strongest imaginable) to assist panelists in rating the samples. Samples were tasted in 10ml volumes at room temperature and labeled with 3-digit blinding codes. Samples were presented in randomized, counterbalanced order within each sample solution (e.g. sucrose, citric acid, etc.).

In order to be selected for participation in testing, panelists needed to correctly rank and rate samples for intensity, with a reasonable number of errors. Approximately 25 people successfully completed this procedure.

Panelists selected in the above procedure were deemed qualified for performing Preliminary Taste Testing procedures. The preliminary taste tests are used to evaluate

new compounds for intensity of basic tastes and off-tastes. A small group of panelists (n=5) taste approximately 5 concentrations of the compound (range typically between 1-100uM, in half-log cycles, e.g. 1, 3, 10, 30, and 100uM) in water or buffer and in a solution of 4% (w/v, 117 mM) sucrose to evaluate enhancement. Typically samples also contain 0.1% ethanol in order to aid dispersion of the compound in a water-based solution. Panelists rate the five basic tastes (sweet, salty, sour, bitter, and umami) as well as off-tastes (such as chemical, metallic, sulfur) on the LMS. Samples are served in 10ml portions at room temperature. The purpose of the test is to determine the highest concentration at which there is no objectionable off-taste, and determine if obvious enhancement of sweet taste exists at any of the concentrations tested.

If the compound is effective and does not have objectionable off-tastes, it is tested with a trained (expert panel) in a larger study.

For example: Five panelists evaluated 1, 3, 10, 30, and 100uM XVI-3 in water and in 4% sucrose solution. All samples with compound were balanced for ethanol at 0.1% (aids in dispersion of compound). Panelists were asked to rate basic tastes and off-tastes using the LMS for each sample tasted. When panelists noted sweetness in any sample, they were asked to taste reference samples of sucrose (2, 4, 6, 8% sucrose) to estimate equivalent sweetness.

A trained (expert) panel was used to further evaluate compounds that had been tested with the preliminary taste test.

Panelists for the trained panel were selected from the larger group of qualifying taste panelists. Panelists were further trained on sweet taste by ranking and rating experiments using sucrose solutions. Panelists completed a series of ranking, rating, and difference from reference tests with sweet solutions. In ranking and rating experiments, panelists evaluated sucrose concentrations (2, 4, 6, 8 % (w/v)) sucrose.

Compounds tested by the trained panel were evaluated in difference from reference experiments. Panelists were given reference samples of various concentrations (2,4,6, or 8 % (w/v) sucrose) and asked to rate samples on a scale of –5 to +5 in terms of difference in sweet taste from the reference (score: –5= much less sweet taste than the reference; 0=same sweet taste as the reference; +5=much more sweet taste than the reference). Test samples were solutions with varying amounts of sucrose and compound. Typically, each session compared the reference sample (labeled as REF) to numerous test samples (labeled with 3-digit blinding codes). Tests

typically included various samples with varying concentrations of sucrose, as well as one blind sample of the reference itself, to evaluate panel accuracy. Compounds were tested against the reference in samples with and without 4% or 6% sucrose. All samples were presented in 10ml volumes at room temperature. Futhermore, to determine the sweetness of the compound alone, a reference solution was prepared at the designated concentration and compared to the threshold sweetness of sucrose (2%).

Example 20

HTS Protocol for Umami Tastants

HEK-$G_{\alpha 15}$ cells were engineered to inducibly express T1R1/T1R3 using the GeneSwitch system (Invitrogen). pGene-derived zeocin-resistant expression vectors for human T1R1 and T1R3 (plasmid SXV603 for T1R1 and SXV611 for T1R3) and a puromycin-resistant pSwitch-derived vector that carries the GeneSwitch protein (plasmid SXV628) were linearized and cotransfected into the HEK-$G_{\alpha 15}$ cell line. Zeocin-resistant and puromycin-resistant colonies were selected, expanded, induced with variable amounts of mifepristone, and tested by calcium imaging for responses to umami taste stimuli. Cells were selected in 0.5 μg/ml puromycin (CAL BIOCHEM) and 100 μg/ml Zeocin (Invitrogen) at 37°C in Dulbecco's Modified Eagle Medium supplemented with GlutaMAX, (10 % dialyzed FBS, and 3 ug/ml blasticidin. Resistant colonies were expanded, and their responses to umami taste stimuli following induction with 10^{-10} M mifepristone determined by fluorescence microscopy following the methods of Li, *et al.*, *PNAS* (2002) 99(7):4692-4696. For automated fluorometric imaging on FLIPR instrumentation (Molecular Device), cells from one clone (designated clone I-17) were seeded into 96- or 384-well plates (approximately 80,000 cell per well) in the presence of 10^{-10} M mifepristone and incubated for 48 hours. Cells were then loaded with the calcium dye fluo-4-AM (Molecular Probes), 3 μM in PBS, for 1.5 hours at room temperature. After replacement with 50 μl PBS, stimulation was performed at room temperature by the addition of 50 μl PBS supplemented with different stimuli. The maximal fluorescence from four independent determinations were averaged, corrected for background fluorescence measured prior to compound addition, and normalized to the response to 0.002 mM ionomycin (CALBIOCHEM).

Example 21

Taste Test Protocol for Umami Tastants

Basic Training of Sensory Tasters: Tasters were trained to evaluate the taste of aqueous solutions (5 mL each, "swash and spit") of the following standard taste compounds by using the triangle test as described in the literature: sucrose (50 mM) for sweet taste; citric acid (5 mM) or lactic acid (20 mM) for sour taste; NaCl (12 mM) for salty taste, quinine (10 μM) or caffeine (1 mM) for bitter taste; and monosodium glutamate (8 mM) for umami or "savory" taste.

Training for Umami Taste: Tasters were given 1-3 sets of 6 MSG and/or MSG-IMP samples ranging from 3-60 mM MSG and 0-200 μM IMP, each arranged in the tray in ascending concentration. This exercise gave the subject practice doing dose response evaluations. Then another set was made up of the same six samples, but were given in random order. The subject was then asked to arrange the samples in ascending intensity and then to rate their umami intensity.

Qualifying Taste Panelists: Tasters were subjected to a standard two alternative forced choice (2AFC) test with 5 pairs of taste samples. They were asked to make a choice of the most umami sample from two samples (a pair). The test contains two easy pairs, two with medium difficulty, and one difficult pair. Tasters who could differentiate the medium difficulty pairs were selected as panelists.

Pilot/Qualitative taste test of Umami Enhancer Candidate (UEC) by a small group of panelists: Taste samples of appropriate concentrations (usually 1-50 μM) were made in water (use minimum amount of ethanol if not soluble); Taste UEC alone at 30 and/or 50 μM for umami and other attributes. Rate those taste attributes on the appropriate Labeled Magnitude Scale (LMS) on the screening ballot; if UEC has no/low umami and other tastes, then move forward to discrimination test; compare certain concentration of MSG, *e.g.*, 12 mM and 12 mM MSG + 30 μM UEC to determine if there is any enhancement; rate the perceived umami intensity on the appropriate LMS on the screening ballot; vary concentration of UEC and/or MSG to find the best combination; decide what solutions to use in panel screening; record all procedures and data including description of study, sample prep, sample arrangement, ballots and sign up sheet for panelists, data entry and evaluation.

2AFC Panel Screening of UEC: Run panel screening with qualified panelists using protocols generated from the pilot tasting; record all procedures and data; prepare summary report with statistically significant conclusions, if any.

Example 22

Quantitative Taste Tests for Compounds 2725761 and 3756807

Quantitative taste tests for compounds 2725761 and 3756807 were run according to procedures presented above. It was found that both of them have some enhancement for MSG, in addition to their additive effect of the umami intensity.

Example 23

Synthesis of Compounds 2725761 and 3756807

Compounds 2725761 and 3756807 are prepared as shown in Example 22, from their corresponding acids and amines. The products are purified by conventional methods, *e.g.*, basic and acidic aqueous washes, or preparative HPLC. The structures of those compounds were confirmed based on usual analytical methods, *e.g.*, NMR and LCMS. This method can also be used to synthesize any of the compounds found in Tables 1-5.

Example 24

Cell Based Assays

Cells were grown and maintained at 37°C in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% FBS and MEM non-essential amino acids (Gibco BRL); media for $G_{\alpha 15}$ cells also contained 3 μg ml^{-1} blasticidin (Gibco BRL). For calcium-imaging experiments, cells were first seeded onto 48-well tissue-culture plates (approximately 30,000 cells per well), and transfected using Mirus TransIt-293 (PanVera). Transfection efficiencies, which were estimated by cotransfection with an RFP expression vector, were typically approximately 60%. To minimize glutamate-induced and glucose-induced desensitization, supplemented DMEM was replaced with low-glucose DMEM supplemented with GlutaMAX and 10% dialyzed FBS (Gibco BRL) approximately 24 hours after transfection. After an additional 24 hours, cells were loaded with the calcium dye fluo-4-AM (Molecular Probes), 3 μM in Dulbecco's PBS buffer (DPBS, GibcoBRL), for 1.5 hours at room temperature. After replacement with 100 μl DPBS, stimulation was performed at room temperature by addition of 100 μl DPBS supplemented with taste stimuli. Calcium mobilization was monitored on an Axiovert S100 microscope equipped with an inverted 10X/0.5 LWD plano fluor objective (Zeiss) and a cooled CCD camera (Princeton Instruments). Fluorescence

images were acquired at 480 nm excitation and 535 nm emission, and analyzed with Imaging Workbench 4.0 software (Axon Instruments). T1R receptor activity was quantitated by counting the number of responding cells 30 seconds after stimulus addition.

What is claimed:

1. A non-naturally occurring compound that specifically binds to a T1R2/T1R3 receptor composed of hT1R2/hT1R3 but not rT1R2/rT1R3.

2. A non-naturally occurring compound that specifically binds to a T1R2/T1R3 receptor composed of hT1R2/rT1R3 but not rT1R2/hT1R3.

3. A non-naturally occurring compound that specifically binds to the N-terminal extracellular domain of T1R2 of the hT1R2/hT1R3 receptor.

4. A non-naturally occurring compound that specifically binds to a T1R2/T1R3 receptor composed of rT1R2/hT1R3 but not hT1R2/rT1R3.

5. A non-naturally occurring compound that specifically binds to hT1R2/hT1R3 and rT1R2/r3-h3 but not to rT1R2/rT1R3 or to hT1R2/h3-r3.

6. A non-naturally occurring compound that specifically binds to hT1R2/hT1R3 and r2-h2/rT1R3 but not to rT1R2/rT1R3 or to h2-r2/hT1R3.

7. A non-naturally occurring compound that specifically binds to the Venus Flytrap Domain (VFD) of T1R2 of the hT1R2/hT1R3 and hT1R2/rT1R3 receptor.

8. A non-naturally occurring compound that specifically binds to amino acid residues 144 and 302 of the human N-terminal Venus flytrap domain of the T1R2 subunit of theT1R2/T1R3 receptor.

9. A non-naturally occurring compound that specifically binds to the N-terminal Venus flytrap domain of the T1R2 subunit of the T1R2/T1R3 receptor, wherein the compound is about 12x5x5 angstroms.

10. A non-naturally occurring compound that specifically binds to the cysteine-rich region of T1R2 of the hT1R2/hT1R3 receptor.

11. A non-naturally occurring compound that specifically binds to the Transmembrane Domain (TM) of T1R2 of the hT1R2/hT1R3 receptor.

12. A non-naturally occurring compound that specifically binds to the human N-terminal extracellular domain of the T1R3 subunit of the T1R2/T1R3.

13. A non-naturally occurring compound that specifically binds to the Venus Flytrap Domain (VFD) of T1R3 of the hT1R2/hT1R3 receptor.

14. A non-naturally occurring compound that specifically binds to the Transmembrane Domain (TM) of T1R3 of the hT1R2/hT1R3 receptor.

15. A non-naturally occurring compound that specifically binds to extracellular loop 2 and extracellular loop 3 of the human transmembrane domain of the T1R3 subunit of T1R2/T1R3.

16. The compound of any one of claims 1-15 that demonstrates compound-dependent increase in fluorescence with an activity compared to the maximal activity for fructose of at least 25% in a fluorescence-based assay using a FLIPR (Molecular Devices) instrument.

17. The compound of any of claims 1-16 that demonstrates a compound-dependent decrease in the EC_{50} for a sweetener by at least two-fold in a fluorescence-based assay using a FLIPR (Molecular Devices) instrument.

18. The compound of any one of claims 1-17 that results in at least 10 out of 100 cells transfected with wild-type or chimeric receptor showing a compound-dependent increase in fluorescence.

19. The compound of any one of claims 1-18 that demonstrates a compound-dependent increase of at least two-fold of the number of fluorescent cells in response to a sub-maximal level of a sweetener.

20. The compound of any one of claims 1-16 that demonstrates a compound-dependent increase in the response of cells to a sub-maximal level of a sweetener of at least 1.25-fold compared to the response to the sweetener alone.

21. The compound of claim 20, wherein the response is measured by fluorescence, calcium levels, IP3 levels, cAMP levels, GTPγS binding, or reporter gene activity (e.g. luciferase, beta-galactosidase).

22. The compound of any one of claims 1-21, having one or more of the following characteristics in a cell:

decreased EC_{50} compared to a control of at 50%,
increased intracellular Ca2+ levels by at least approximately 25%,
increased intracellular cAMP by at least approximately 25%,
increased intracellular cGMP by at least approximately 25%,
increased intracellular IP_3 by at least approximately 25%, or
increased G protein binding of GTPγS by at least approximately 25%.

23. A chimeric T1R2/T1R3 receptor comprising, a human T1R2 subunit and a rat T1R3 subunit.

24. A chimeric T1R2/T1R3 receptor comprising, a rat T1R2 subunit and a human T1R3 subunit.

25. A chimeric T1R2 receptor subunit comprising, a human extracellular domain, a rat transmembrane domain and a rat intracellular domain.

26. A chimeric T1R3 receptor subunit comprising, a rat extracellular domain, a human transmembrane domain and a human intracellular domain.

27. A non-naturally occurring compound that binds to the N-terminal extracellular domain of T1R1 of the T1R1/hT1R3 receptor.

28. A non-naturally occurring compound that binds to the T1R1VFD of the T1R1/T1R3 savory receptor.

29. A non-naturally occurring compound that binds to the cysteine-rich region of T1R1 of the T1R1/hT1R3 receptor.

30. A non-naturally occurring compound that binds to the T1R1 TM domain of the T1R1/T1R3 savory receptor.

31. A non-naturally occurring compound that binds to the N-terminal extracellular domain of T1R3 of the T1R1/hT1R3 receptor.

32. A non-naturally occurring compound that binds to the T1R3 VFD of the T1R1/T1R3 savory receptor.

33. A non-naturally occurring compound that binds to the cysteine-rich region of T1R3 of the T1R1/hT1R3 receptor.

34. A non-naturally occurring compound that binds to the T1R3 TM domain of the T1R1/T1R3 savory receptor.

35. A non-naturally occurring compound that binds to the TM domain of T1R1 of a truncated savory receptor composed of the h1TM/h3TM.

36. A non-naturally occurring compound that binds to the TM domain of T1R3 of a truncated sweet receptor composed of h1TM/h3TM.

37. A non-naturally occurring compound that binds to the TM domain of T1R1 of a chimeric receptor composed of mGluR-h1/mGluR-h3.

38. A non-naturally occurring compound that binds to the TM domain of T1R3 of a chimeric receptor composed of mGluR-h1/mGluR-h3.

39. The compound of any one of claims 27-38 that demonstrates compound-dependent increase in fluorescence with an activity compared to the maximal activity of glutamate of at least 25% in a fluorescence-based assay using a FLIPR (Molecular Devices) instrument.

40. The compound of any one of claims 27-38 that demonstrates a compound-dependent decrease in the EC50 for glutamate by at least two-fold in a fluorescence-based assay using a FLIPR (Molecular Devices) instrument.

41. The compound of any one of claims 27-38 that results in at least 10 out of 100 transfected cells showing a compound-dependent increase in fluorescence measured with a fluorescent microscope.

42. The compound of any one of claims 27-38 that results in a compound-dependent increase of at least two-fold of the number of fluorescent cells in response to a sub-maximal level of glutamate.

43. The compound of any one of claims 27-38 that results in a compound-dependent increase in the response of cells to a sub-maximal level of glutamate of at least 1.25-fold compared to the response to glutamate alone.

44. The compound of claim 43, wherein the response is measured by fluorescence, calcium levels, IP3 levels, cAMP levels, ΓΤΠγS binding, or reporter gene activity.

45. A method for identifying compounds that modulate taste perception by identifying compounds that bind to, activate, inhibit, enhance and/or modulate one or more of the receptors of any one of claims 23-26.

46. A method for identifying a compound that modulates sweet taste perception comprising comparing the effect of the compound on a sweet receptor to the effect of a compound of any one of claims 1-22, an enhancement of sweet perception approximately equal to or greater than the sweet enhancement of the compound indicating a compound that enhances sweet perception.

47. A method for identifying a compound that modulates umami taste perception comprising comparing the effect of the compound on an umami receptor to the effect of a compound of any one of claims 27-44, an enhancement of savory perception approximately equal to or greater than the savory enhancement of the compound indicating a compound that enhances umami perception.

48. A method for identifying compounds that modulate taste perception by identifying compounds that bind to, activate, inhibit, and/or modulate a receptor expressed by a cell that stably expresses one or more of the receptors of any one of claims 23-26.

49. A method for modulating the savory taste of a comestible or medicinal product comprising:

providing at least one comestible or medicinal product, or a precursor thereof, and

combining the comestible or medicinal product or precursor thereof with at least a savory flavor modulating amount of at least one non-naturally occurring compound of any one of claims 27-44, or a comestibly acceptable salt thereof, so as to form a modified comestible or medicinal product;

thereby modulating the savory taste of a comestible or medicinal product.

50. A method for inhibiting the savory taste of a comestible or medicinal product comprising:

providing at least one comestible or medicinal product, or a precursor thereof, and

combining the comestible or medicinal product or precursor thereof with at least a savory flavor inhibiting amount of at least one non-naturally occurring compound of any one of claims 27-44, or a comestibly acceptable salt thereof, so as to form a modified comestible or medicinal product;

thereby inhibiting the savory taste of a comestible or medicinal product.

51. A method for increasing the savory taste of a comestible or medicinal product comprising:

providing at least one comestible or medicinal product, or a precursor thereof, and

combining the comestible or medicinal product or precursor thereof with at least a savory flavor increasing amount of at least one non-naturally occurring compound of any one of claims 27-44, or a comestibly acceptable salt thereof, so as to form a modified comestible or medicinal product;

thereby increasing the savory taste of a comestible or medicinal product.

52. A method for modulating the sweet taste of a comestible or medicinal product comprising:

providing at least one comestible or medicinal product, or a precursor thereof, and

combining the comestible or medicinal product or precursor thereof with at least a sweet flavor modulating amount of at least one non-naturally occurring compound of any one of claims 1-22 and 60-61, or a comestibly acceptable salt thereof, so as to form a modified comestible or medicinal product;

thereby modulating the sweet taste of a comestible or medicinal product.

53. A method for inhibiting the sweet taste of a comestible or medicinal product comprising:

providing at least one comestible or medicinal product, or a precursor thereof, and

combining the comestible or medicinal product or precursor thereof with at least a sweet flavor inhibiting amount of at least one non-naturally occurring compound of any one of claims 1-22 and 60-61, or a comestibly acceptable salt thereof, so as to form a modified comestible or medicinal product;

thereby inhibiting the sweet taste of a comestible or medicinal product.

54. A method for increasing the sweet taste of a comestible or medicinal product comprising:

providing at least one comestible or medicinal product, or a precursor thereof, and

combining the comestible or medicinal product or precursor thereof with at least a sweet flavor increasing amount of at least one non-naturally occurring compound of any one of claims 1-22 or 60-61, or a comestibly acceptable salt thereof, so as to form a modified comestible or medicinal product;

thereby increasing the sweet taste of a comestible or medicinal product.

55. A method of enhancing umami taste perception comprising contacting an umami receptor with cyclamate and NHDC, and their derivatives.

56. A method of enhancing umami taste perception comprising contacting an umami receptor with lactisole derivatives.

57. A method of enhancing sweet taste perception comprising contacting a sweet receptor with cyclamate and NHDC, and their derivatives.

58. A method of enhancing sweet taste perception comprising contacting a sweet receptor with lactisole derivatives.

59. The compound of any one of claims 27-44, wherein the compound is not sucrose, fructose, glucose, erythritol, isomalt, lactitol, mannitol, sorbitol, xylitol, certain known natural terpenoids, flavonoids, or protein sweeteners, di-peptides,

tri-peptides aspartame, saccharin, sucralose, halogenated saccharides, acesulfame-K, cyclamate, sucralose, alitame, monosodium glutamate ("MSG"), inosine monophosphate (IMP), adenosine monophosphate, or guanosine monophosphate (GMP).

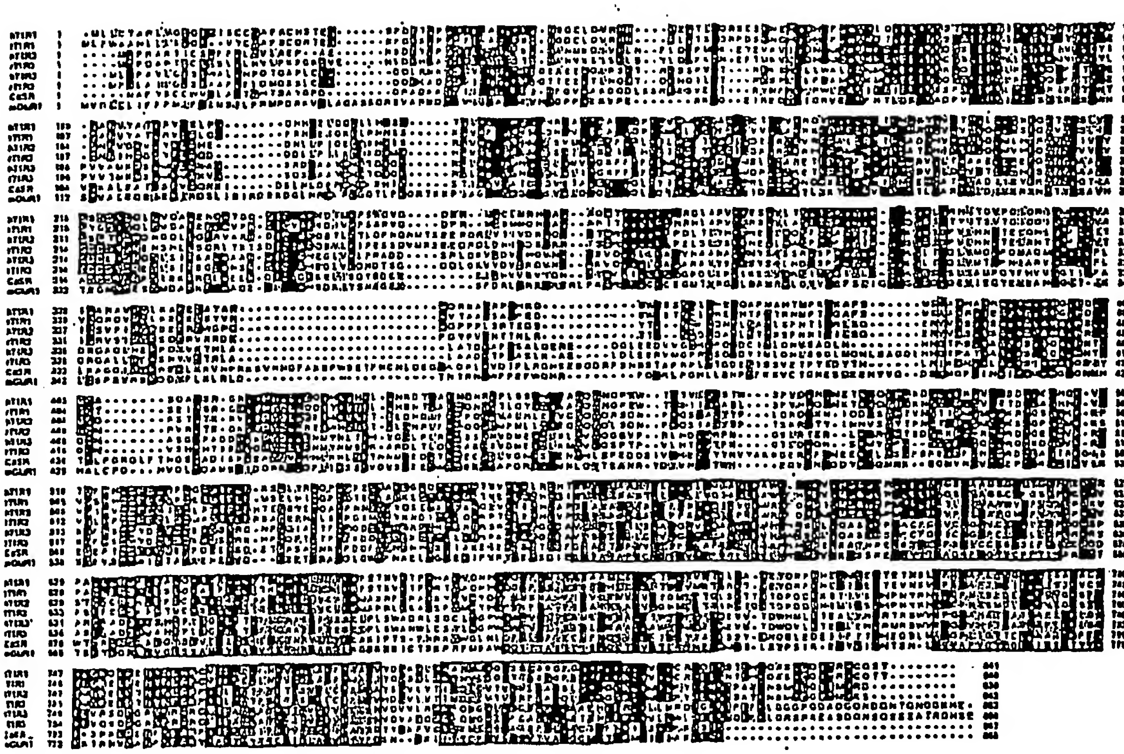
60. The compound of any one of claims 1-22, wherein the compound is not sucrose, fructose, glucose, erythritol, isomalt, lactitol, mannitol, sorbitol, xylitol, certain known natural terpenoids, flavonoids, or protein sweeteners, di-peptides, tri-peptides, aspartame, saccharin, sucralose, halogenated saccharides, acesulfame-K, cyclamate, sucralose, and alitame. neotame, perillartine, SC-45647, SC-40014, monellin, NC-002740-01, thaumatin, CC-00100, NC-00420, alitame, SC-44102, dulcin, NC-00576, slycyrrhizic Acid, stevioside, Na-Saccharin, D-tryptophan, cyclamate, DHB, glycolic Acid, glycine, D (-)fructose, homofuronol, D (-) tagatose, maltose, D (+) glucose, D-sorbitol, D (+) galactose, α-lactose, L()fructose, L (+), compound 403249, or glucose.

61. The compound of any one of claims 1-22, wherein the compound is not sucrose, fructose, glucose, erythritol, isomalt, lactitol, mannitol, sorbitol, xylitol, certain known natural terpenoids, flavonoids, or protein sweeteners, di-peptides, tri-peptides, aspartame, saccharin, sucralose, halogenated saccharides, acesulfame-K, cyclamate, sucralose, and alitame. neotame, perillartine, SC-45647, SC-40014, monellin, NC-002740-01, thaumatin, CC-00100, NC-00420, alitame, SC-44102, dulcin, NC-00576, slycyrrhizic Acid, stevioside, Na-Saccharin, D-tryptophan, cyclamate, DHB, glycolic Acid, glycine, D (-)fructose, homofuronol, D (-) tagatose, maltose, D (+) glucose, D-sorbitol, D (+) galactose, α-lactose, L()fructose, L (+), compound 403249, glucose, or Compound 6364395.

62. The compound of any one of claims 27-44, wherein the compound is not sucrose, fructose, glucose, erythritol, isomalt, lactitol, mannitol, sorbitol, xylitol, certain known natural terpenoids, flavonoids, or protein sweeteners, di-peptides, tri-peptides aspartame, saccharin, sucralose, halogenated saccharides, acesulfame-K, cyclamate, sucralose, alitame, monosodium glutamate ("MSG"),

inosine monophosphate (IMP), adenosine monophosphate, or Compound 6364395, guanosine monophosphate (GMP).

Figure 1: Catalog of human and rat T1Rs



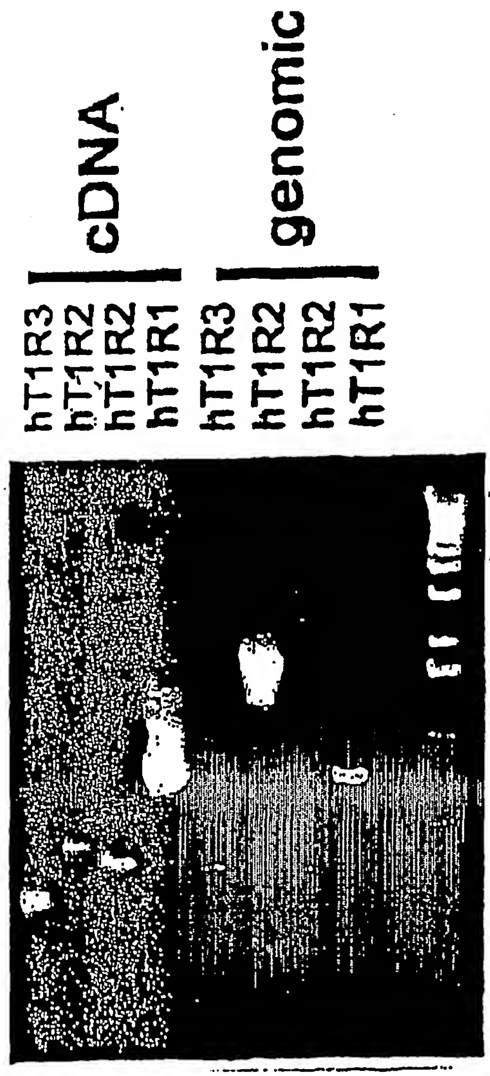


Figure 2 hT1R2 and hT1R3 are expressed in human tongue epithelium. cDNA-specific amplification products can be amplified from cDNA prepared from resected human circumvallate papillae.

Figure 3 Human T1R2/T1R3 functions as a sweet taste receptor

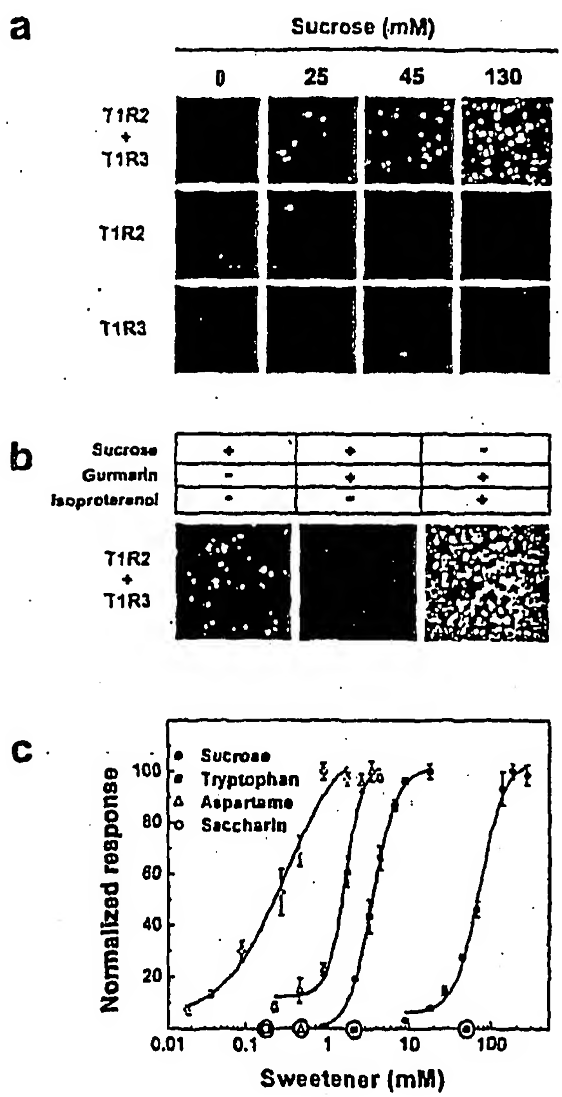
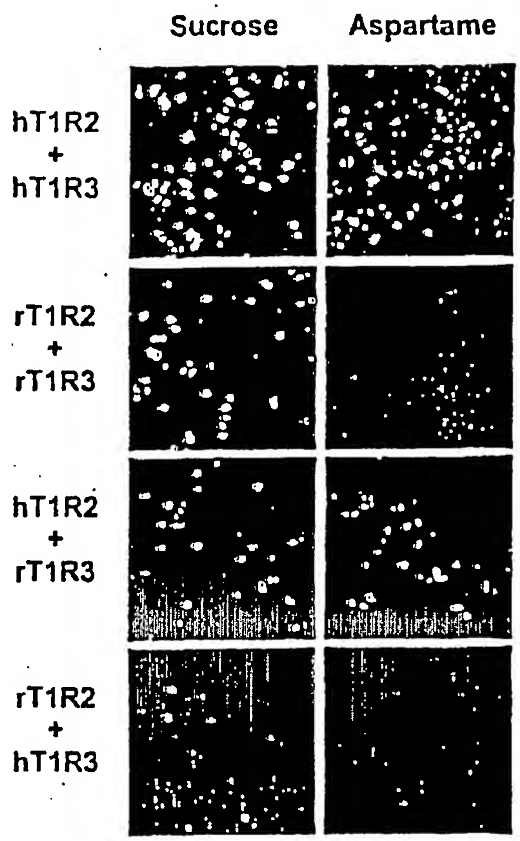


Figure 4 T1R2 may control T1R2/T1R3 ligand specificity



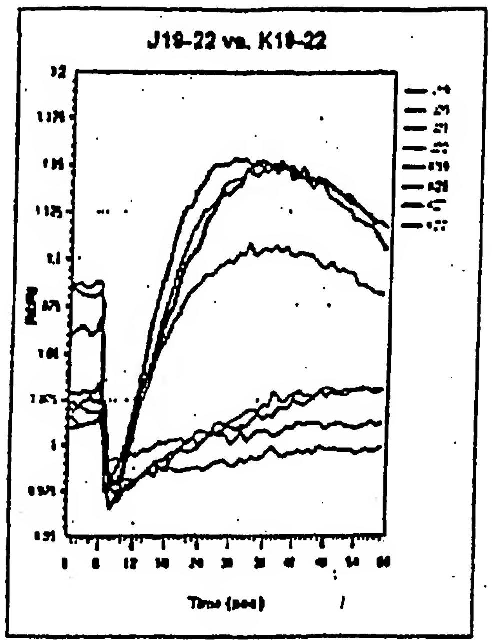


Figure 5

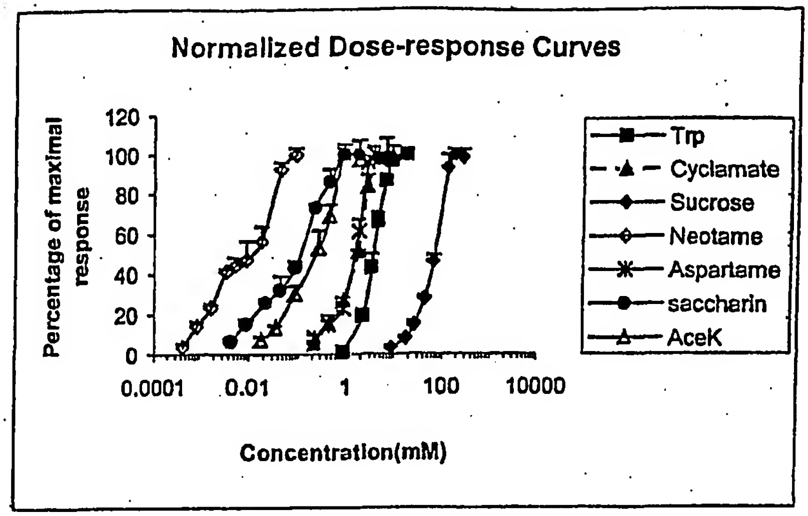
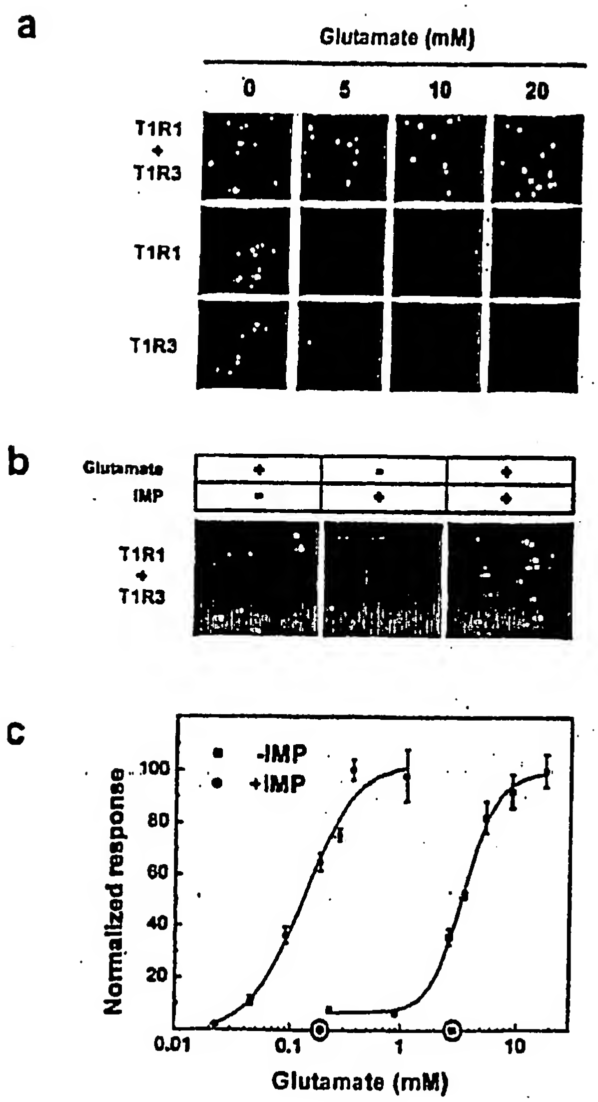

Normalized Dose-response Curves
Percentage of maximal response
120
100
80
60
40
20
0
0.0001
0.01
1
100
10000
Concentration(mM)
Trp
Cyclamate
Sucrose
Neotame
Aspartame
saccharin
AceK

Figure 6

Figure 7 Key ligand-binding residues of mGlurR1 are conserved in T1R1

Figure 8 Human T1R1/T1R3 functions as an umami taste receptor



A. B.

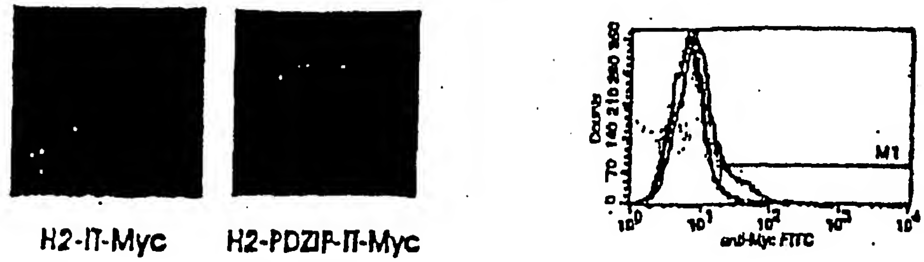


Figure 9 PDZIP facilitate the surface expression of human T1R2.

A. Immunofluorescence staining of Myc-tagged hT1R2 indicates that PDZIP significantly increases the amount of human T1R2 protein on the plasma membrane.

B. FACS analysis data demonstrating the same result. Myc-tagged human T1R2: Green line. Myc-tagged

C. human T1R2 with PDZIP: black line.

Figure 10 Calcium-imaging data demonstrating hT1R2/hT1R3 responses to a number of sweet stimuli.

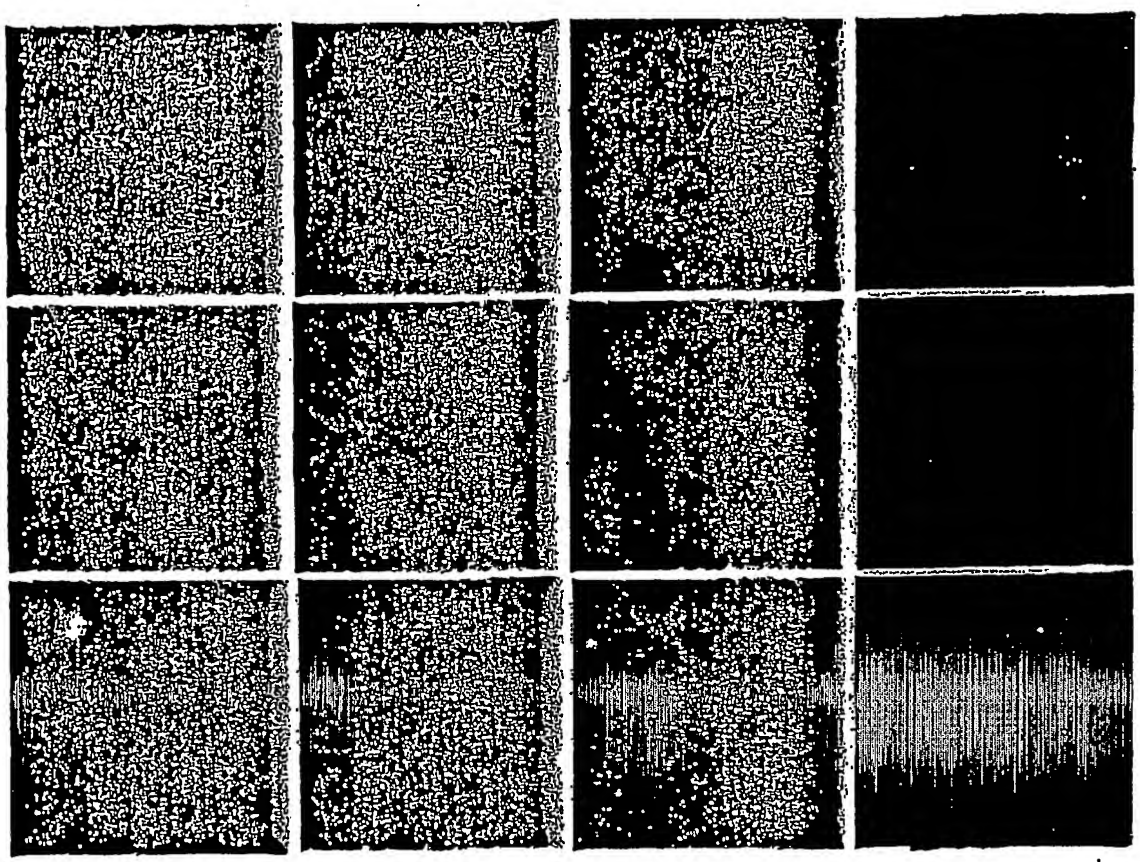


Figure 11

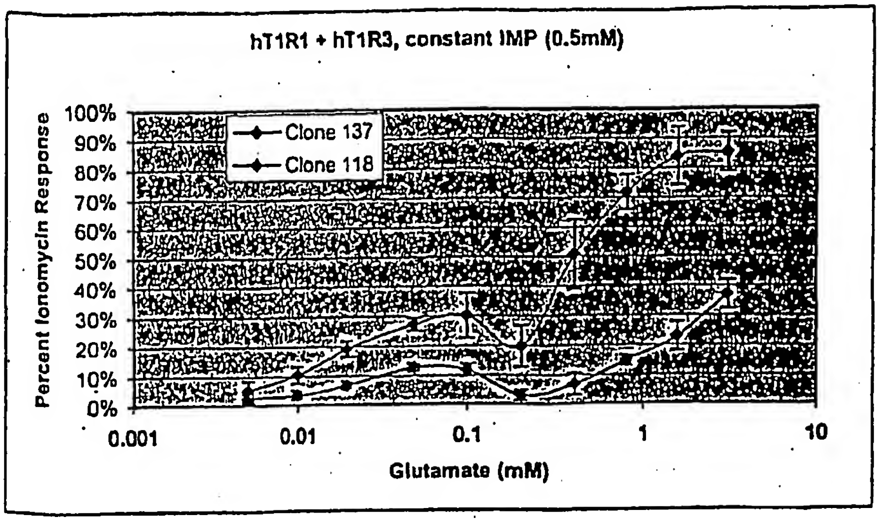


Figure 12

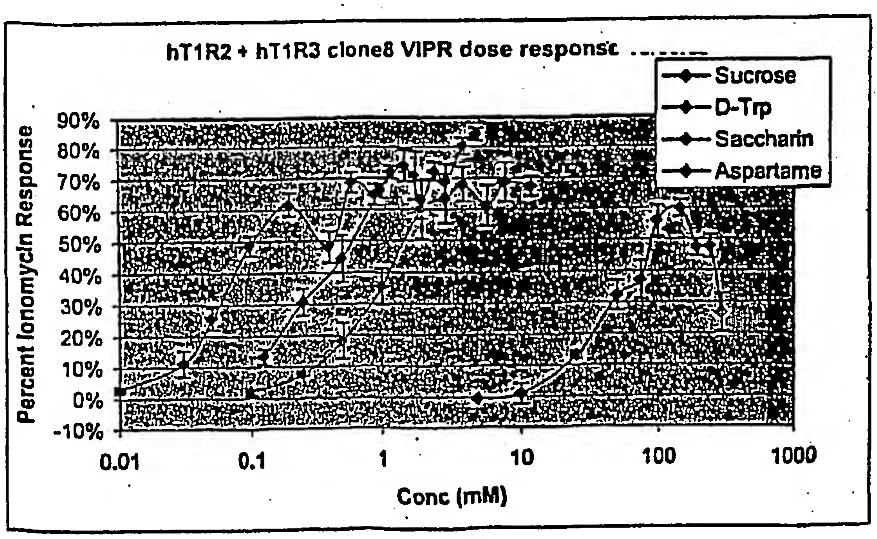
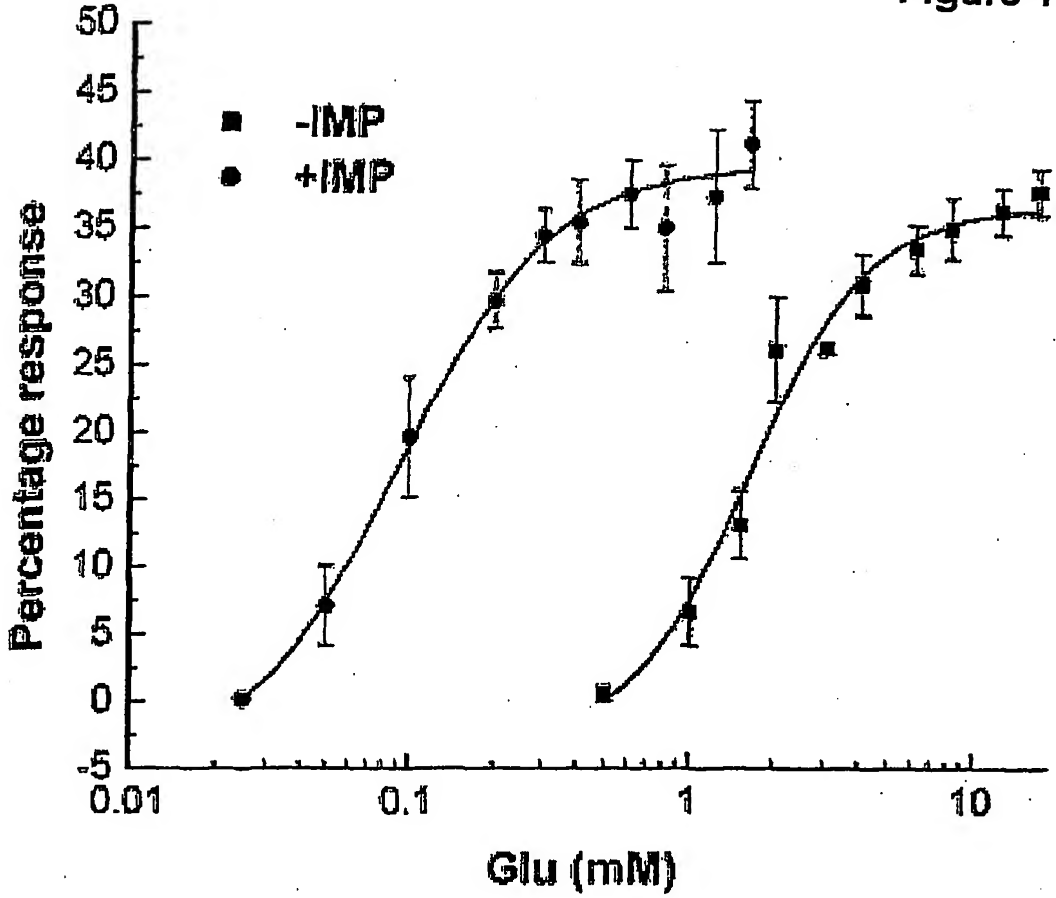


Figure 13


-IMP
+IMP
Percentage response
50
45
40
35
30
25
20
15
10
5
0
-5
0.01
0.1
1
10
Glu (mM)

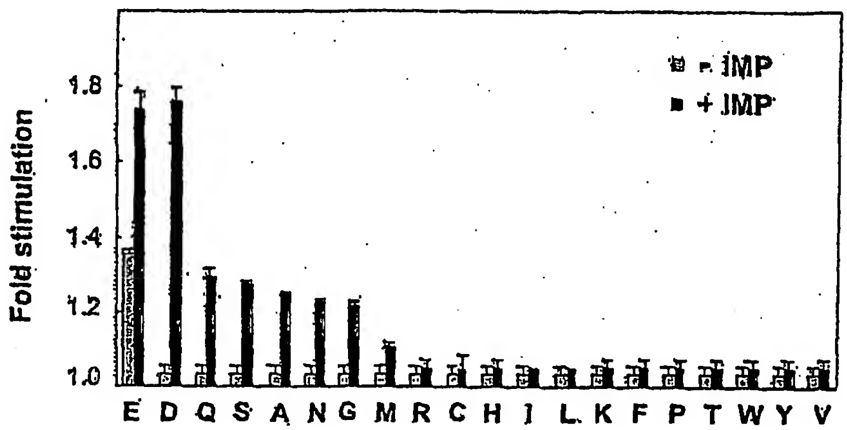

- IMP
+ IMP
Fold stimulation
1.8
1.6
1.4
1.2
1.0
E D Q S A N G M R C H I L K F P T W Y V

Figure 14

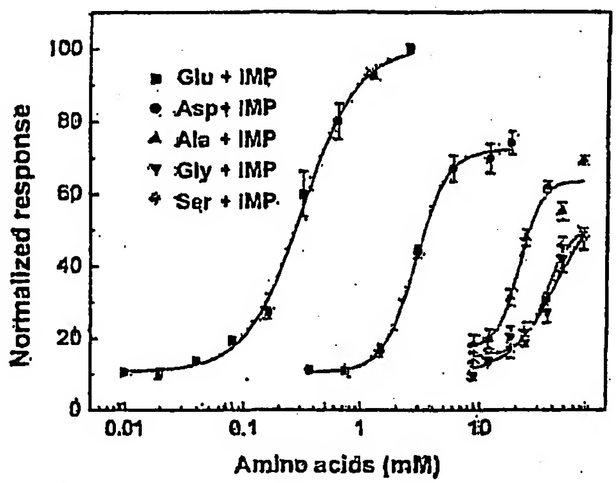

Glu + IMP
Asp + IMP
Ala + IMP
Gly + IMP
Ser + IMP
Normalized response
0
20
40
60
80
100
0.01
0.1
1
10
Amino acids (mM)

Figure 15

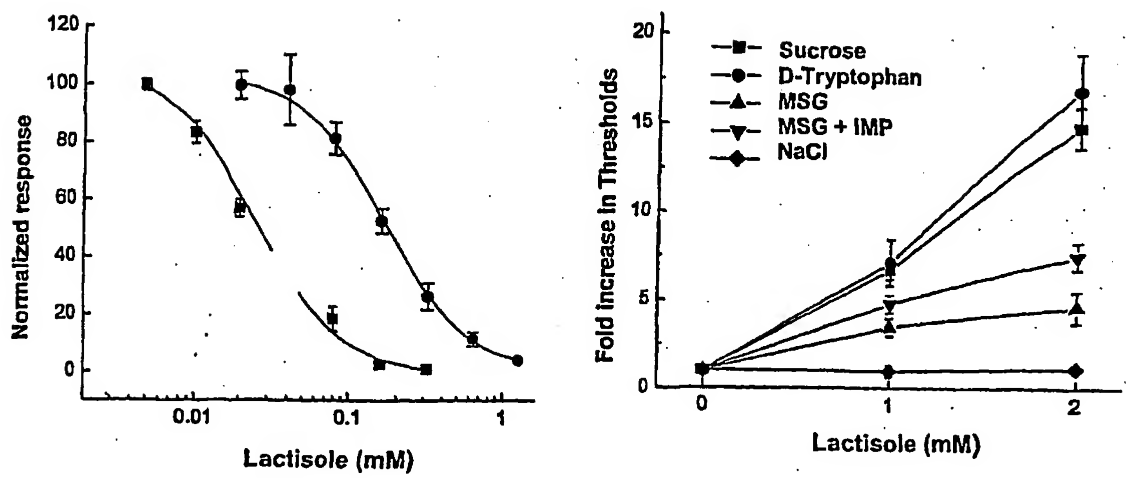


Figure 16 Lactisole inhibits the T1R2/T1R3 sweet and T1R1/T1R3 umami receptors and sweet and umami taste. (*Left panel*) responses of HEK-$G_{\alpha 15}$ cells transiently transfected with T1R1/T1R3 (*circles*) to 10 mM L-glutamate and HEK-$G_{\alpha 15}$ cells transiently transfected with T1R2/T1R3 (*squares*) to 150 mM sucrose in the presence of variable concentrations of lactisole are shown. (*Right panel*) fold increases in taste detection thresholds in the presence of 1 and 2 mM lactisole are shown for the sweet taste stimuli sucrose and D-tryptophan, the umami taste stimuli L-glutamate (MSG) and L-glutamate plus 0.2 mM IMP, and sodium chloride. Detection thresholds were determined following the method of Schiffman et al.

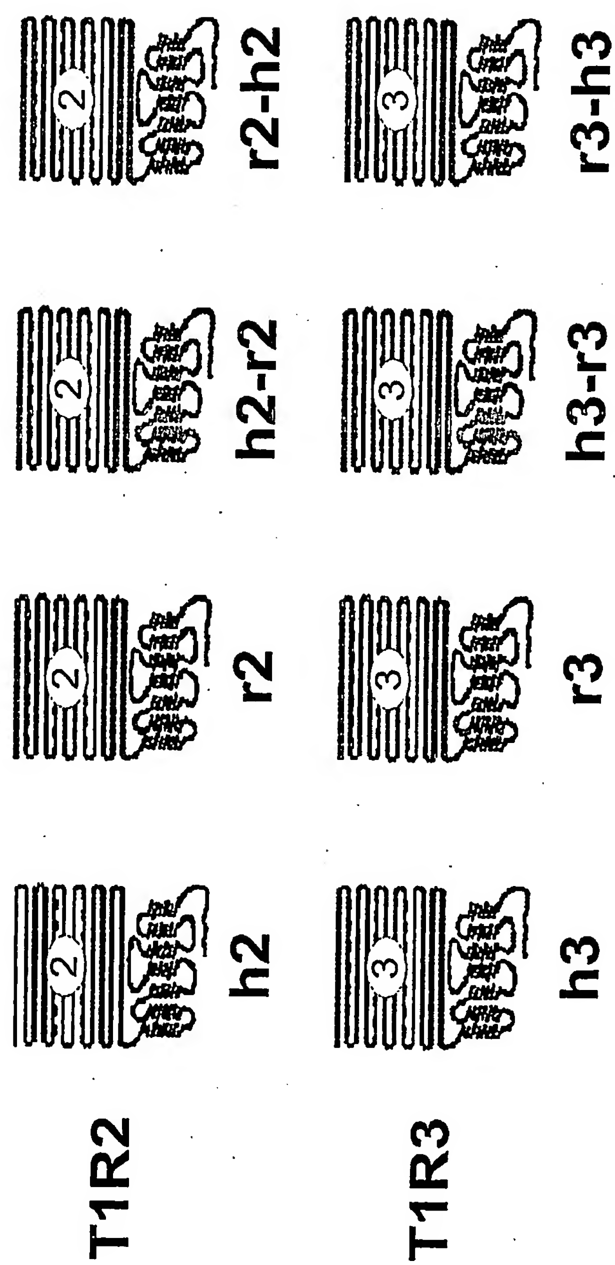
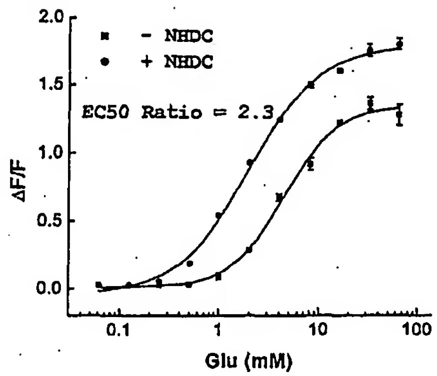


Fig. 17

Figure 18


2.0
1.5
1.0
0.5
0.0
- NHDC
+ NHDC
EC50 Ratio = 2.3
ΔF/F
0.1
1
10
100
Glu (mM)

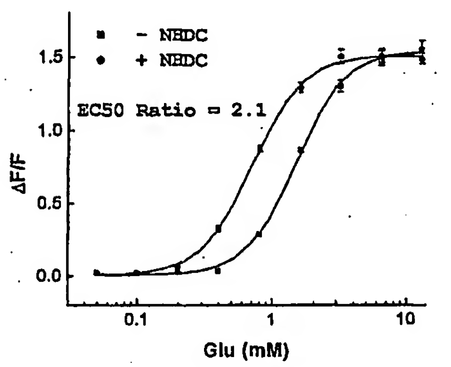
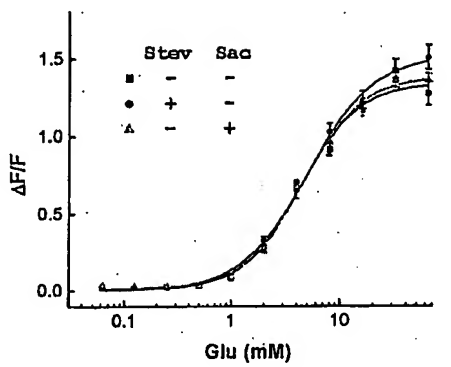

1.5
1.0
0.5
0.0
- NHDC
+ NHDC
EC50 Ratio = 2.1
ΔF/F
0.1
1
10
Glu (mM)

Figure 19


Stev Sac
■ – –
● + –
△ – +
ΔF/F
1.5
1.0
0.5
0.0
0.1
1
10
Glu (mM)

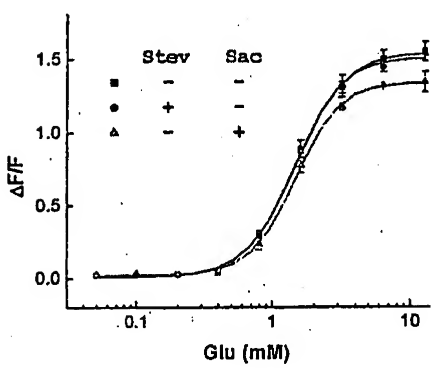

Stev Sac
■ – –
● + –
△ – +
ΔF/F
1.5
1.0
0.5
0.0
0.1
1
10
Glu (mM)

Figure 20

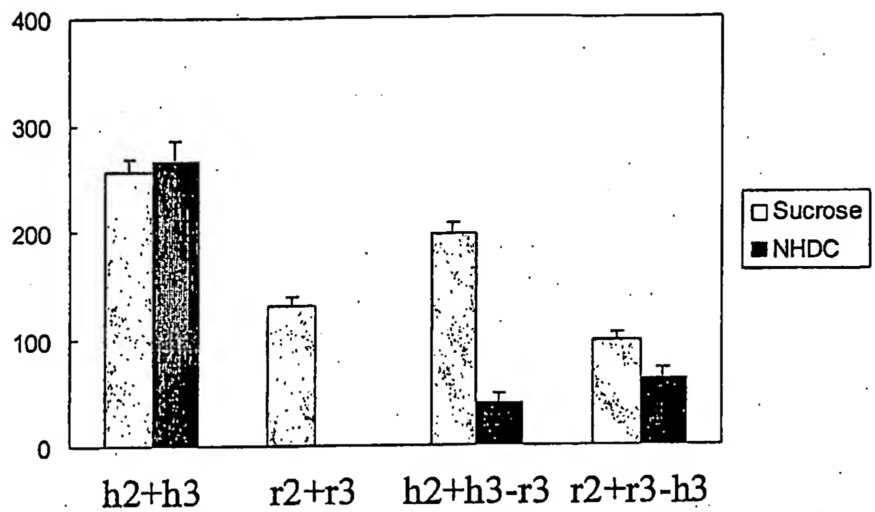


Figure 21A

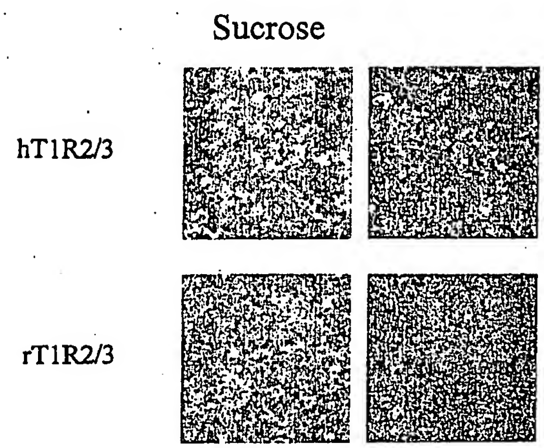
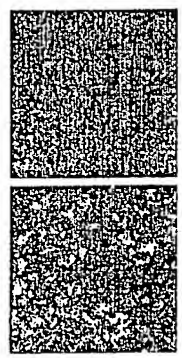


Figure 21B

h2-r2 + h3

r2-h2 + r3



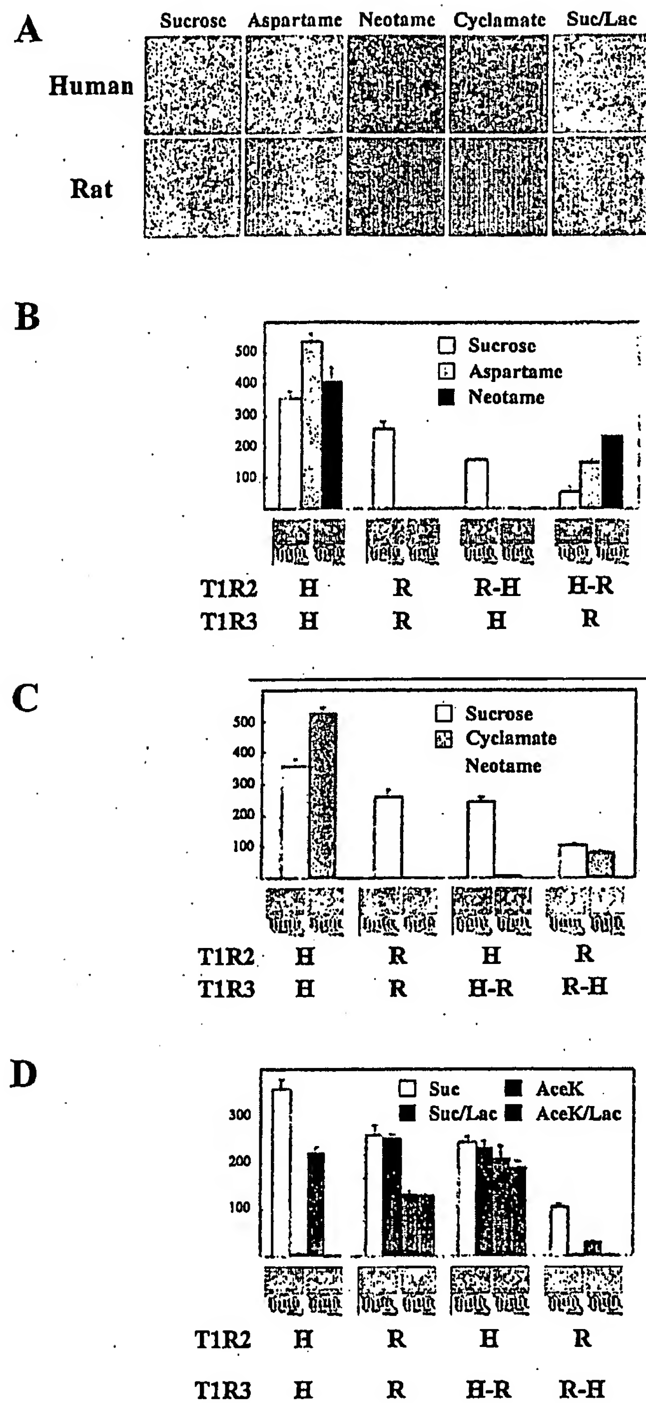
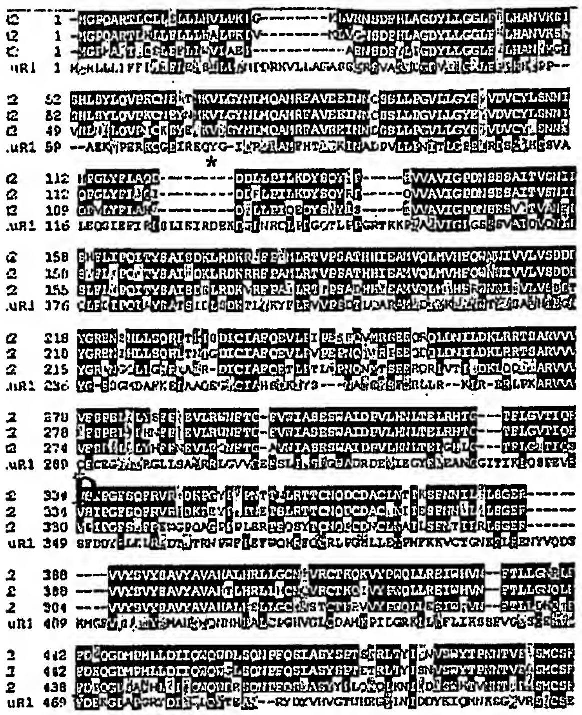
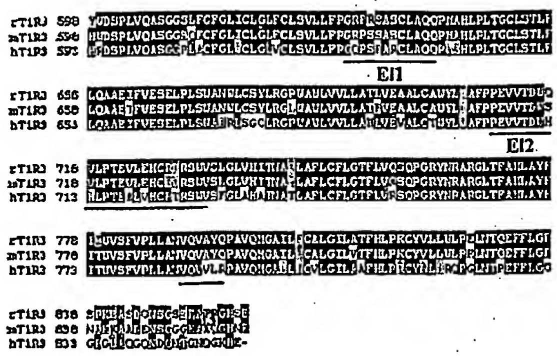
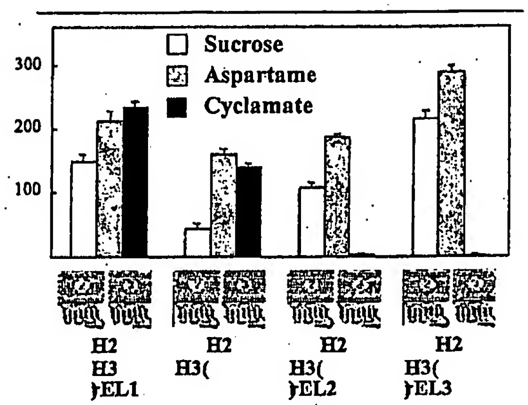

A
Sucrose
Aspartame
Neotame
Cyclamate
Suc/Lac
Human
Rat
B
500
400
300
200
100
Sucrose
Aspartame
Neotame
T1R2 H R R-H H-R
T1R3 H R H R
C
500
400
300
200
100
Sucrose
Cyclamate
Neotame
T1R2 H R H R
T1R3 H R H-R R-H
D
300
200
100
Suc
AceK
Suc/Lac
AceK/Lac
T1R2 H R H R
T1R3 H R H-R R-H

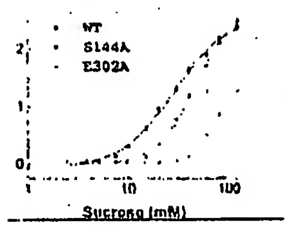
Fig. 22

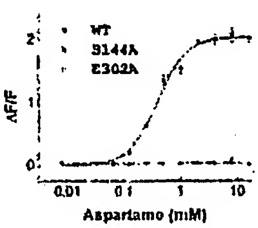


C


rT1R3
mT1R3
hT1R3
El1
El2


Sucrose
Aspartame
Cyclamate
300
200
100
H2
H3
}EL1
H2
H3(
H2
H3(
}EL2
H2
H3(
}EL3


WT
S144A
E302A
Sucrose (mM)


WT
S144A
E302A
ΔF/F
0.01
0.1
1
10
Aspartame (mM)

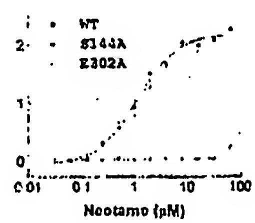

WT
S144A
E302A
0.01
0.1
1
10
100
Neotame (µM)

Fig. 23

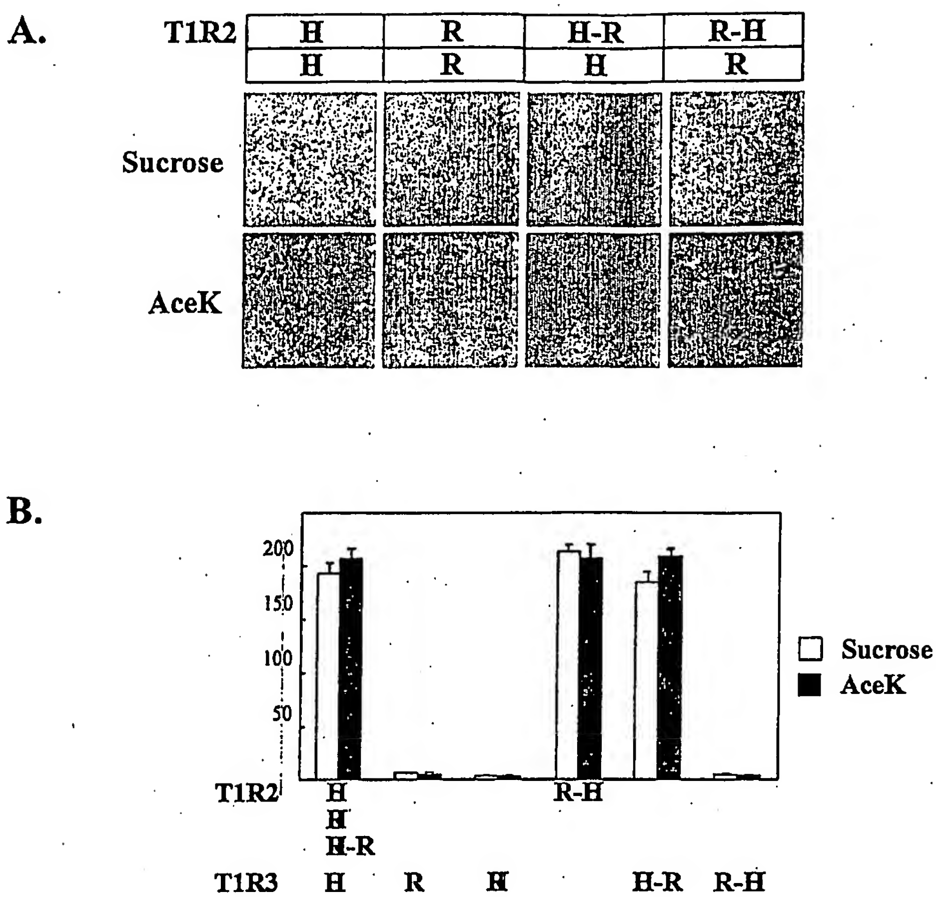

A.
T1R2
H
R
H-R
R-H
H
R
H
R
Sucrose
AceK
B.
200
150
100
50
Sucrose
AceK
T1R2
H
H
H-R
R-H
T1R3
H
R
H
H-R
R-H

Fig. 24

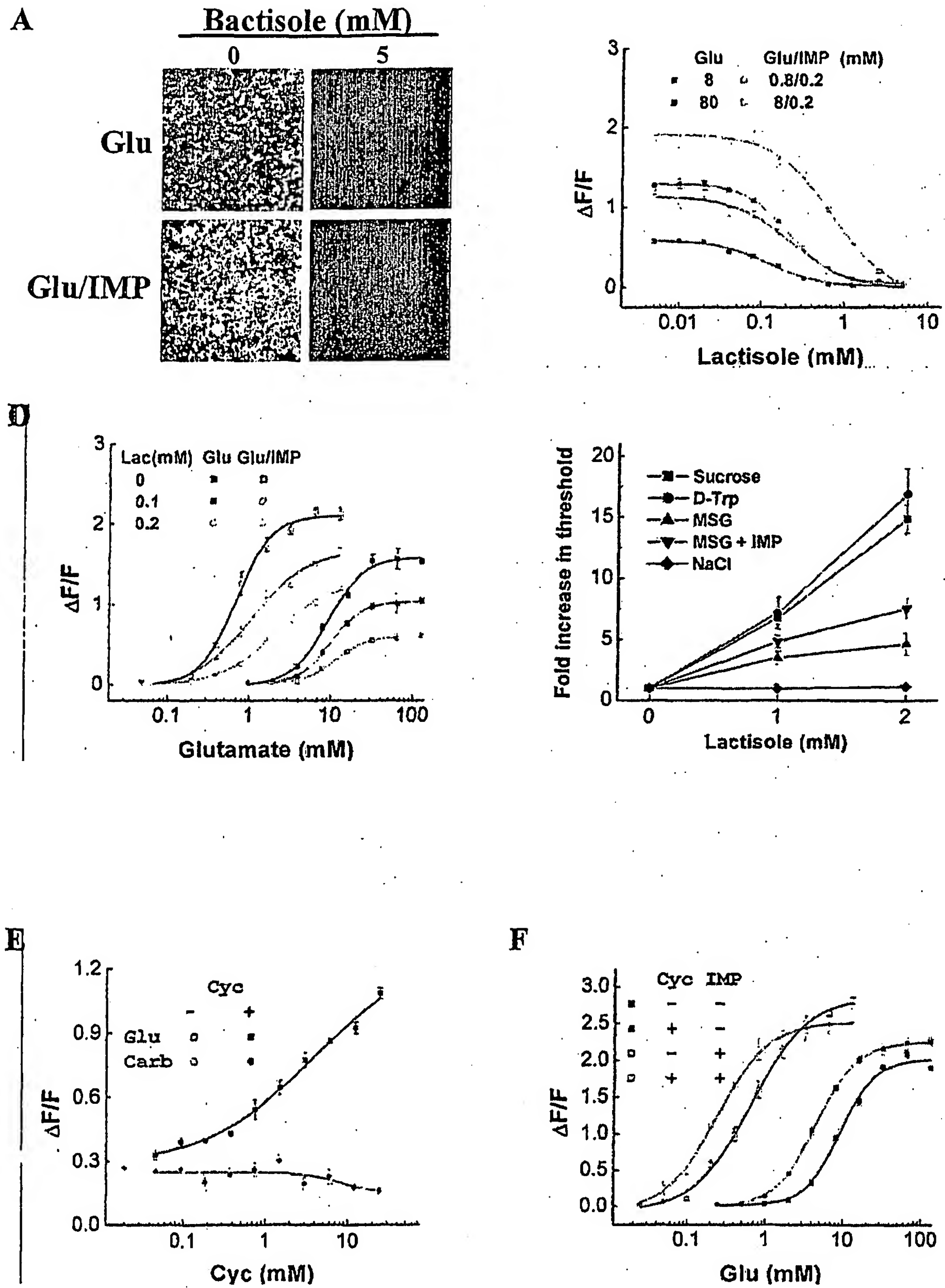

A
Bactisole (mM)
0
5
Glu
Glu/IMP
Glu
Glu/IMP (mM)
8
0.8/0.2
80
8/0.2
ΔF/F
Lactisole (mM)
D
Lac(mM) Glu Glu/IMP
0
0.1
0.2
ΔF/F
Glutamate (mM)
Sucrose
D-Trp
MSG
MSG + IMP
NaCl
Fold increase in threshold
Lactisole (mM)
E
Cyc
Glu
Carb
ΔF/F
Cyc (mM)
F
Cyc IMP
ΔF/F
Glu (mM)

Fig. 25

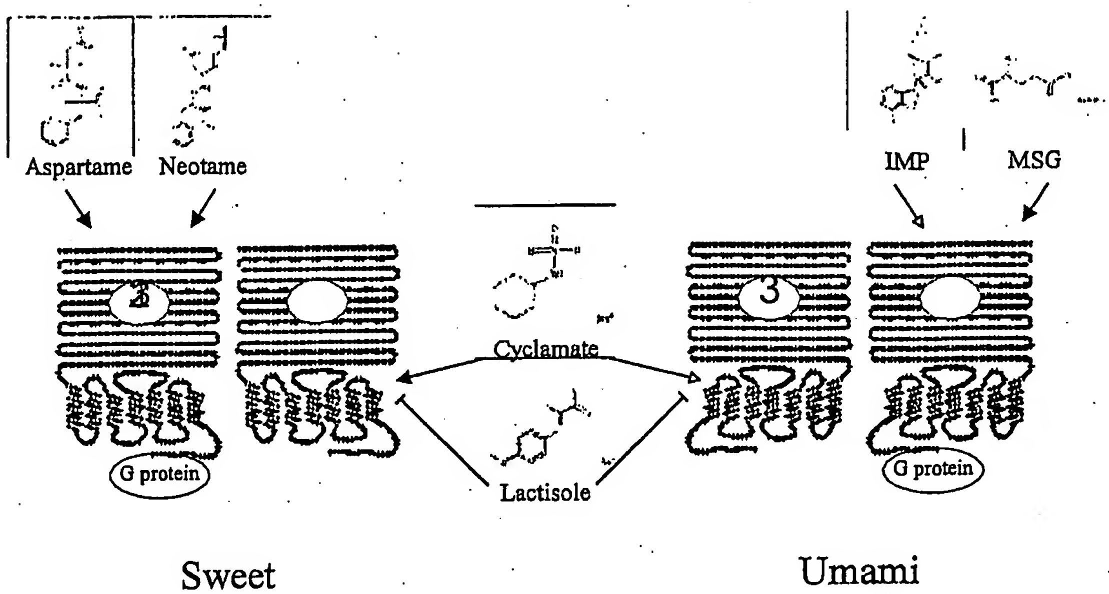
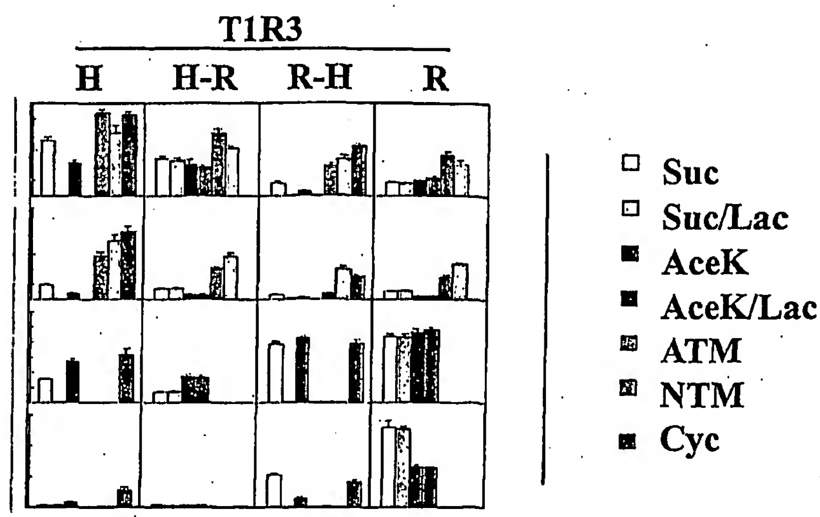

Aspartame
Neotame
IMP
MSG
3
3
Cyclamate
Lactisole
G protein
G protein
Sweet
Umami

Fig. 26

A


T1R3
H
H-R
R-H
R
Suc
Suc/Lac
AceK
AceK/Lac
ATM
NTM
Cyc

B

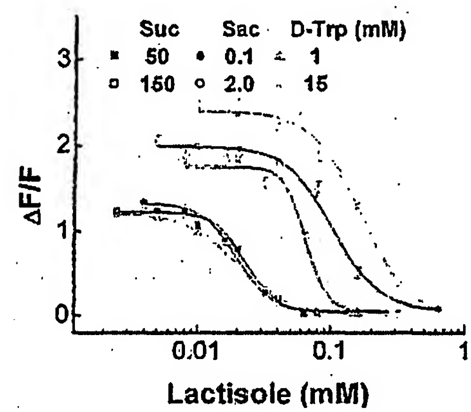
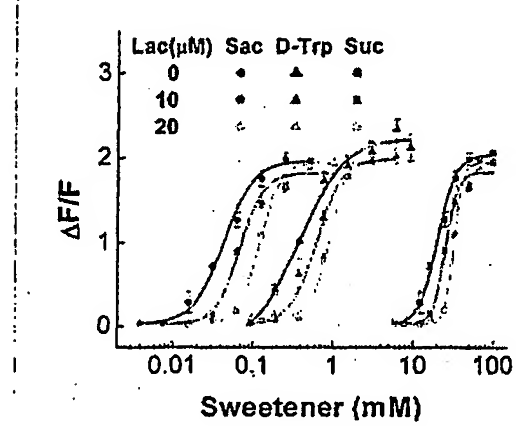

Suc Sac D-Trp (mM)
50 0.1 1
150 2.0 15
ΔF/F
Lactisole (mM)

Fig. 27

C


Lac(μM) Sac D-Trp Suc
0
10
20
ΔF/F
Sweetener (mM)

SEQUENCE LISTING

<110> Senomyx, Inc.

<120> T1R HETERO-OLIGOMERIC TASTE RECEPTORS,
CELL LINES THAT EXPRESS SAID RECEPTORS, AND TASTE COMPOUNDS

<130> 19328.0001u1

<140> Unassigned
<141> 2004-08-06

<150> 60/494,071
<151> 2003-08-06

<150> 60/552,064
<151> 2004-03-09

<160> 33

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence; note =
synthetic construct

<400> 1
Ser Val Ser Thr Trp
1 5

<210> 2
<211> 14
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence; note =
synthetic construct

<221> VARIANT
<222> 1,3,4,6,7,8,11,12,13
<223> At position 1, Xaa can be either Thr or Arg
At position 3, Xaa can be either Phe or Leu
At position 4, Xaa can be either Arg, Gln, or Pro
At position 6, Xaa can be either Arg or Thr
At position 7, Xaa can be either Ser, Pro or Val
At position 8, Xaa can be either Val, Glu, Arg,
Lys or Thr
At position 11, Xaa can be either Ala or Glu
At position 12, Xaa can be either Trp or Leu

At position 13, Xaa can be either Arg, His or Gly

<400> 2
Xaa Cys Xaa Xaa Arg Xaa Xaa Xaa Phe Leu Xaa Xaa Xaa Glu
1 5 10

<210> 3
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence; note =
synthetic construct

<221> VARIANT
<222> 1,3,4,7,9,10,11,13,14,15
<223> At position 1, Xaa can be either Leu or Gln
At position 3, Xaa can be either Glu, Gly or Thr
At position 4, Xaa can be either Asn, Arg or Cys
At position 7, Xaa can be either Arg or Glu
At position 9, Xaa can be either Arg or Lys
At position 10, Xaa can be either Cys, Gly or Phe
At position 11, Xaa can be either Val, Leu or Ile
At position 13, Xaa can be either Phe or Leu
At position 14, Xaa can be either Ala or Ser
At position 15, Xaa can be either Met or Leu

<400> 3
Xaa Pro Xaa Xaa Tyr Asn Xaa Ala Xaa Xaa Xaa Thr Xaa Xaa Xaa
1 5 10 15

<210> 4
<211> 858
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence; note =
synthetic construct

<400> 4
Met Pro Gly Leu Ala Ile Leu Gly Leu Ser Leu Ala Ala Phe Leu Glu
1 5 10 15
Leu Gly Met Gly Ser Ser Leu Cys Leu Ser Gln Gln Phe Lys Ala Gln
20 25 30
Gly Asp Tyr Ile Leu Gly Gly Leu Phe Pro Leu Gly Thr Thr Glu Glu
35 40 45
Ala Thr Leu Asn Gln Arg Thr Gln Pro Asn Gly Ile Leu Cys Thr Arg
50 55 60
Phe Ser Pro Leu Gly Leu Phe Leu Ala Met Ala Met Lys Met Ala Val
65 70 75 80
Glu Glu Ile Asn Asn Gly Ser Ala Leu Leu Pro Gly Leu Arg Leu Gly
85 90 95
Tyr Asp Leu Phe Asp Thr Cys Ser Glu Pro Val Val Thr Met Lys Pro
100 105 110
Ser Leu Met Phe Met Ala Lys Val Gly Ser Gln Ser Ile Ala Ala Tyr

```
        115                 120                 125
Cys Asn Tyr Thr Gln Tyr Gln Pro Arg Val Leu Ala Val Ile Gly Pro
    130                 135                 140
His Ser Ser Glu Leu Ala Leu Ile Thr Gly Lys Phe Phe Ser Phe Phe
145                 150                 155                 160
Leu Met Pro Gln Val Ser Tyr Ser Ala Ser Met Asp Arg Leu Ser Asp
                165                 170                 175
Arg Glu Thr Phe Pro Ser Phe Phe Arg Thr Val Pro Ser Asp Arg Val
            180                 185                 190
Gln Leu Gln Ala Val Val Thr Leu Leu Gln Asn Phe Ser Trp Asn Trp
        195                 200                 205
Val Ala Ala Leu Gly Ser Asp Asp Asp Tyr Gly Arg Glu Gly Leu Ser
    210                 215                 220
Ile Phe Ser Gly Leu Ala Asn Ser Arg Gly Ile Cys Ile Ala His Glu
225                 230                 235                 240
Gly Leu Val Pro Gln His Asp Thr Ser Gly Gln Gln Leu Gly Lys Val
                245                 250                 255
Val Asp Val Leu Arg Gln Val Asn Gln Ser Lys Val Gln Val Val Val
            260                 265                 270
Leu Phe Ala Ser Ala Arg Ala Val Tyr Ser Leu Phe Ser Tyr Ser Ile
        275                 280                 285
Leu His Asp Leu Ser Pro Lys Val Trp Val Ala Ser Glu Ser Trp Leu
    290                 295                 300
Thr Ser Asp Leu Val Met Thr Leu Pro Asn Ile Ala Arg Val Gly Thr
305                 310                 315                 320
Val Leu Gly Phe Leu Gln Arg Gly Ala Leu Leu Pro Glu Phe Ser His
                325                 330                 335
Tyr Val Glu Thr Arg Leu Ala Leu Ala Ala Asp Pro Thr Phe Cys Ala
            340                 345                 350
Ser Leu Lys Ala Glu Leu Asp Leu Glu Glu Arg Val Met Gly Pro Arg
        355                 360                 365
Cys Ser Gln Cys Asp Tyr Ile Met Leu Gln Asn Leu Ser Ser Gly Leu
    370                 375                 380
Met Gln Asn Leu Ser Ala Gly Gln Leu His His Gln Ile Phe Ala Thr
385                 390                 395                 400
Tyr Ala Ala Val Tyr Ser Val Ala Gln Ala Leu His Asn Thr Leu Gln
                405                 410                 415
Cys Asn Val Ser His Cys His Thr Ser Glu Pro Val Gln Pro Trp Gln
            420                 425                 430
Leu Leu Glu Asn Met Tyr Asn Met Ser Phe Arg Ala Arg Asp Leu Thr
        435                 440                 445
Leu Gln Phe Asp Ala Lys Gly Ser Val Asp Met Glu Tyr Asp Leu Lys
    450                 455                 460
Met Trp Val Trp Gln Ser Pro Thr Pro Val Leu His Thr Val Gly Thr
465                 470                 475                 480
Phe Asn Gly Thr Leu Gln Leu Gln His Ser Lys Met Tyr Trp Pro Gly
                485                 490                 495
Asn Gln Val Pro Val Ser Gln Cys Ser Arg Gln Cys Lys Asp Gly Gln
            500                 505                 510
Val Arg Arg Val Lys Gly Phe His Ser Cys Cys Tyr Asp Cys Val Asp
        515                 520                 525
Cys Lys Ala Gly Ser Tyr Arg Lys His Pro Asp Asp Phe Thr Cys Thr
    530                 535                 540
Pro Cys Gly Lys Asp Gln Trp Ser Pro Glu Lys Ser Thr Thr Cys Leu
545                 550                 555                 560
Pro Arg Arg Pro Lys Phe Leu Ala Trp Gly Glu Pro Ala Val Leu Ser
                565                 570                 575
```

Leu Leu Leu Leu Leu Cys Leu Val Leu Gly Leu Thr Leu Ala Ala Leu
580 585 590
Gly Leu Phe Val His Tyr Trp Asp Ser Pro Leu Val Gln Ala Ser Gly
595 600 605
Gly Ser Leu Phe Cys Phe Gly Leu Ile Cys Leu Gly Leu Phe Cys Leu
610 615 620
Ser Val Leu Leu Phe Pro Gly Arg Pro Arg Ser Ala Ser Cys Leu Ala
625 630 635 640
Gln Gln Pro Met Ala His Leu Pro Leu Thr Gly Cys Leu Ser Thr Leu
645 650 655
Phe Leu Gln Ala Ala Glu Ile Phe Val Glu Ser Glu Leu Pro Leu Ser
660 665 670
Trp Ala Asn Trp Leu Cys Ser Tyr Leu Arg Gly Pro Trp Ala Trp Leu
675 680 685
Val Val Leu Leu Ala Thr Leu Val Glu Ala Ala Leu Cys Ala Trp Tyr
690 695 700
Leu Met Ala Phe Pro Pro Glu Val Val Thr Asp Trp Gln Val Leu Pro
705 710 715 720
Thr Glu Val Leu Glu His Cys Arg Met Arg Ser Trp Val Ser Leu Gly
725 730 735
Leu Val His Ile Thr Asn Ala Val Leu Ala Phe Leu Cys Phe Leu Gly
740 745 750
Thr Phe Leu Val Gln Ser Gln Pro Gly Arg Tyr Asn Arg Ala Arg Gly
755 760 765
Leu Thr Phe Ala Met Leu Ala Tyr Phe Ile Ile Trp Val Ser Phe Val
770 775 780
Pro Leu Leu Ala Asn Val Gln Val Ala Tyr Gln Pro Ala Val Gln Met
785 790 795 800
Gly Ala Ile Leu Phe Cys Ala Leu Gly Ile Leu Ala Thr Phe His Leu
805 810 815
Pro Lys Cys Tyr Val Leu Leu Trp Leu Pro Glu Leu Asn Thr Gln Glu
820 825 830
Phe Phe Leu Gly Arg Ser Pro Lys Glu Ala Ser Asp Gly Asn Ser Gly
835 840 845
Ser Ser Glu Ala Thr Arg Gly His Ser Glu
850 855

<210> 5
<211> 841
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence; note =
synthetic construct

<400> 5
Met Leu Leu Cys Thr Ala Arg Leu Val Gly Leu Gln Leu Leu Ile Ser
1 5 10 15
Cys Cys Trp Ala Phe Ala Cys His Ser Thr Glu Ser Ser Pro Asp Phe
20 25 30
Thr Leu Pro Gly Asp Tyr Leu Leu Ala Gly Leu Phe Pro Leu His Ser
35 40 45
Gly Cys Leu Gln Val Arg His Arg Pro Glu Val Thr Leu Cys Asp Arg
50 55 60
Ser Cys Ser Phe Asn Glu His Gly Tyr His Leu Phe Gln Ala Met Arg
65 70 75 80

Leu Gly Val Glu Glu Ile Asn Asn Ser Thr Ala Leu Leu Pro Asn Ile
85 90 95
Thr Leu Gly Tyr Gln Leu Tyr Asp Val Cys Ser Asp Ser Ala Asn Val
100 105 110
Tyr Ala Thr Leu Arg Val Leu Ser Leu Pro Gly Gln His His Ile Glu
115 120 125
Leu Gln Gly Asp Leu Leu His Tyr Ser Pro Thr Val Leu Ala Val Ile
130 135 140
Gly Pro Asp Ser Thr Asn Arg Ala Ala Thr Thr Ala Ala Leu Leu Ser
145 150 155 160
Pro Phe Leu Val Pro Met Ile Ser Tyr Ala Ala Ser Ser Glu Thr Leu
165 170 175
Ser Val Lys Arg Gln Tyr Pro Ser Phe Leu Arg Thr Ile Pro Asn Asp
180 185 190
Lys Tyr Gln Val Glu Thr Met Val Leu Leu Leu Gln Lys Phe Gly Trp
195 200 205
Thr Trp Ile Ser Leu Val Gly Ser Ser Asp Asp Tyr Gly Gln Leu Gly
210 215 220
Val Gln Ala Leu Glu Asn Gln Ala Thr Gly Gln Gly Ile Cys Ile Ala
225 230 235 240
Phe Lys Asp Ile Met Pro Phe Ser Ala Gln Val Gly Asp Glu Arg Met
245 250 255
Gln Cys Leu Met Arg His Leu Ala Gln Ala Gly Ala Thr Val Val Val
260 265 270
Val Phe Ser Ser Arg Gln Leu Ala Arg Val Phe Phe Glu Ser Val Val
275 280 285
Leu Thr Asn Leu Thr Gly Lys Val Trp Val Ala Ser Glu Ala Trp Ala
290 295 300
Leu Ser Arg His Ile Thr Gly Val Pro Gly Ile Gln Arg Ile Gly Met
305 310 315 320
Val Leu Gly Val Ala Ile Gln Lys Arg Ala Val Pro Gly Leu Lys Ala
325 330 335
Phe Glu Glu Ala Tyr Ala Arg Ala Asp Lys Lys Ala Pro Arg Pro Cys
340 345 350
His Lys Gly Ser Trp Cys Ser Ser Asn Gln Leu Cys Arg Glu Cys Gln
355 360 365
Ala Phe Met Ala His Thr Met Pro Lys Leu Lys Ala Phe Ser Met Ser
370 375 380
Ser Ala Tyr Asn Ala Tyr Arg Ala Val Tyr Ala Val Ala His Gly Leu
385 390 395 400
His Gln Leu Leu Gly Cys Ala Ser Gly Ala Cys Ser Arg Gly Arg Val
405 410 415
Tyr Pro Trp Gln Leu Leu Glu Gln Ile His Lys Val His Phe Leu Leu
420 425 430
His Lys Asp Thr Val Ala Phe Asn Asp Asn Arg Asp Pro Leu Ser Ser
435 440 445

Tyr Asn Ile Ile Ala Trp Asp Trp Asn Gly Pro Lys Trp Thr Phe Thr
450 455 460
Val Leu Gly Ser Ser Thr Trp Ser Pro Val Gln Leu Asn Ile Asn Glu
465 470 475 480
Thr Lys Ile Gln Trp His Gly Lys Asp Asn Gln Val Pro Lys Ser Val
485 490 495
Cys Ser Ser Asp Cys Leu Glu Gly His Gln Arg Val Val Thr Gly Phe
500 505 510
His His Cys Cys Phe Glu Cys Val Pro Cys Gly Ala Gly Thr Phe Leu
515 520 525

Asn Lys Ser Asp Leu Tyr Arg Cys Gln Pro Cys Gly Lys Glu Glu Trp
    530   535   540
Ala Pro Glu Gly Ser Gln Thr Cys Phe Pro Arg Thr Val Val Phe Leu
545   550   555   560
Ala Leu Arg Glu His Thr Ser Trp Val Leu Leu Ala Ala Asn Thr Leu
                565   570   575
Leu Leu Leu Leu Leu Leu Gly Thr Ala Gly Leu Phe Ala Trp His Leu
            580   585   590
Asp Thr Pro Val Val Arg Ser Ala Gly Gly Arg Leu Cys Phe Leu Met
        595   600   605
Leu Gly Ser Leu Ala Ala Gly Ser Gly Ser Leu Tyr Gly Phe Phe Gly
    610   615   620
Glu Pro Thr Arg Pro Ala Cys Leu Leu Arg Gln Ala Leu Phe Ala Leu
625   630   635   640
Gly Phe Thr Ile Phe Leu Ser Cys Leu Thr Val Arg Ser Phe Gln Leu
                645   650   655
Ile Ile Ile Phe Lys Phe Ser Thr Lys Val Pro Thr Phe Tyr His Ala
            660   665   670
Trp Val Gln Asn His Gly Ala Gly Leu Phe Val Met Ile Ser Ser Ala
        675   680   685
Ala Gln Leu Leu Ile Cys Leu Thr Trp Leu Val Val Trp Thr Pro Leu
    690   695   700
Pro Ala Arg Glu Tyr Gln Arg Phe Pro His Leu Val Met Leu Glu Cys
705   710   715   720
Thr Glu Thr Asn Ser Leu Gly Phe Ile Leu Ala Phe Leu Tyr Asn Gly
                725   730   735
Leu Leu Ser Ile Ser Ala Phe Ala Cys Ser Tyr Leu Gly Lys Asp Leu
            740   745   750
Pro Glu Asn Tyr Asn Glu Ala Lys Cys Val Thr Phe Ser Leu Leu Phe
        755   760   765
Asn Phe Val Ser Trp Ile Ala Phe Phe Thr Thr Ala Ser Val Tyr Asp
    770   775   780
Gly Lys Tyr Leu Pro Ala Ala Asn Met Met Ala Gly Leu Ser Ser Leu
785   790   795   800
Ser Ser Gly Phe Gly Gly Tyr Phe Leu Pro Lys Cys Tyr Val Ile Leu
                805   810   815
Cys Arg Pro Asp Leu Asn Ser Thr Glu His Phe Gln Ala Ser Ile Gln
            820   825   830
Asp Tyr Thr Arg Arg Cys Gly Ser Thr
        835   840

<210> 6
<211> 839
<212> PRT
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence; note =
      synthetic construct

<400> 6
Met Gly Pro Arg Ala Lys Thr Ile Cys Ser Leu Phe Phe Leu Leu Trp
1   5   10   15
Val Leu Ala Glu Pro Ala Glu Asn Ser Asp Phe Tyr Leu Pro Gly Asp
            20   25   30
Tyr Leu Leu Gly Gly Leu Phe Ser Leu His Ala Asn Met Lys Gly Ile

```
            35                  40                  45
Val His Leu Asn Phe Leu Gln Val Pro Met Cys Lys Glu Tyr Glu Val
        50                  55                  60
Lys Val Ile Gly Tyr Asn Leu Met Gln Ala Met Arg Phe Ala Val Glu
65                  70                  75                  80
Glu Ile Asn Asn Asp Ser Ser Leu Leu Pro Gly Val Leu Leu Gly Tyr
                85                  90                  95
Glu Ile Val Asp Val Cys Tyr Ile Ser Asn Asn Val Gln Pro Val Leu
            100                 105                 110
Tyr Phe Leu Ala His Glu Asp Asn Leu Leu Pro Ile Gln Glu Asp Tyr
        115                 120                 125
Ser Asn Tyr Ile Ser Arg Val Val Ala Val Ile Gly Pro Asp Asn Ser
    130                 135                 140
Glu Ser Val Met Thr Val Ala Asn Phe Leu Ser Leu Phe Leu Leu Pro
145                 150                 155                 160
Gln Ile Thr Tyr Ser Ala Ile Ser Asp Glu Leu Arg Asp Lys Val Arg
                165                 170                 175
Phe Pro Ala Leu Leu Arg Thr Thr Pro Ser Ala Asp His His Val Glu
            180                 185                 190
Ala Met Val Gln Leu Met Leu His Phe Arg Trp Asn Trp Ile Ile Val
        195                 200                 205
Leu Val Ser Ser Asp Thr Tyr Gly Arg Asp Asn Gly Gln Leu Leu Gly
    210                 215                 220
Glu Arg Val Ala Arg Arg Asp Ile Cys Ile Ala Phe Gln Glu Thr Leu
225                 230                 235                 240
Pro Thr Leu Gln Pro Asn Gln Asn Met Thr Ser Glu Glu Arg Gln Arg
                245                 250                 255
Leu Val Thr Ile Val Asp Lys Leu Gln Gln Ser Thr Ala Arg Val Val
            260                 265                 270
Val Val Phe Ser Pro Asp Leu Thr Leu Tyr His Phe Phe Asn Glu Val
        275                 280                 285
Leu Arg Gln Asn Phe Thr Gly Ala Val Trp Ile Ala Ser Glu Ser Trp
    290                 295                 300
Ala Ile Asp Pro Val Leu His Asn Leu Thr Glu Leu Gly His Leu Gly
305                 310                 315                 320
Thr Phe Leu Gly Ile Thr Ile Gln Ser Val Pro Ile Pro Gly Phe Ser
                325                 330                 335
Glu Phe Arg Glu Trp Gly Pro Gln Ala Gly Pro Pro Pro Leu Ser Arg
            340                 345                 350
Thr Ser Gln Ser Tyr Thr Cys Asn Gln Glu Cys Asp Asn Cys Leu Asn
        355                 360                 365
Ala Thr Leu Ser Phe Asn Thr Ile Leu Arg Leu Ser Gly Glu Arg Val
    370                 375                 380
Val Tyr Ser Val Tyr Ser Ala Val Tyr Ala Val Ala His Ala Leu His
385                 390                 395                 400
Ser Leu Leu Gly Cys Asp Lys Ser Thr Cys Thr Lys Arg Val Val Tyr
                405                 410                 415
Pro Trp Gln Leu Leu Glu Glu Ile Trp Lys Val Asn Phe Thr Leu Leu
            420                 425                 430
Asp His Gln Ile Phe Phe Asp Pro Gln Gly Asp Val Ala Leu His Leu
        435                 440                 445
Glu Ile Val Gln Trp Gln Trp Asp Arg Ser Gln Asn Pro Phe Gln Ser
    450                 455                 460
Val Ala Ser Tyr Tyr Pro Leu Gln Arg Gln Leu Lys Asn Ile Gln Asp
465                 470                 475                 480
Ile Ser Trp His Thr Val Asn Asn Thr Ile Pro Met Ser Met Cys Ser
                485                 490                 495
```

Lys Arg Cys Gln Ser Gly Gln Lys Lys Lys Pro Val Gly Ile His Val
500 505 510
Cys Cys Phe Glu Cys Ile Asp Cys Leu Pro Gly Thr Phe Leu Asn His
515 520 525
Thr Glu Asp Glu Tyr Glu Cys Gln Ala Cys Pro Asn Asn Glu Trp Ser
530 535 540
Tyr Gln Ser Glu Thr Ser Cys Phe Lys Arg Gln Leu Val Phe Leu Glu
545 550 555 560

Trp His Glu Ala Pro Thr Ile Ala Val Ala Leu Leu Ala Ala Leu Gly
565 570 575
Phe Leu Ser Thr Leu Ala Ile Leu Val Ile Phe Trp Arg His Phe Gln
580 585 590
Thr Pro Ile Val Arg Ser Ala Gly Gly Pro Met Cys Phe Leu Met Leu
595 600 605
Thr Leu Leu Leu Val Ala Tyr Met Val Val Pro Val Tyr Val Gly Pro
610 615 620
Pro Lys Val Ser Thr Cys Leu Cys Arg Gln Ala Leu Phe Pro Leu Cys
625 630 635 640
Phe Thr Ile Cys Ile Ser Cys Ile Ala Val Arg Ser Phe Gln Ile Val
645 650 655
Cys Ala Phe Lys Met Ala Ser Arg Phe Pro Arg Ala Tyr Ser Tyr Trp
660 665 670
Val Arg Tyr Gln Gly Pro Tyr Val Ser Met Ala Phe Ile Thr Val Leu
675 680 685
Lys Met Val Ile Val Val Ile Gly Met Leu Ala Thr Gly Leu Ser Pro
690 695 700
Thr Thr Arg Thr Asp Pro Asp Asp Pro Lys Ile Thr Ile Val Ser Cys
705 710 715 720
Asn Pro Asn Tyr Arg Asn Ser Leu Leu Phe Asn Thr Ser Leu Asp Leu
725 730 735
Leu Leu Ser Val Val Gly Phe Ser Phe Ala Tyr Met Gly Lys Glu Leu
740 745 750
Pro Thr Asn Tyr Asn Glu Ala Lys Phe Ile Thr Leu Ser Met Thr Phe
755 760 765
Tyr Phe Thr Ser Ser Val Ser Leu Cys Thr Phe Met Ser Ala Tyr Ser
770 775 780
Gly Val Leu Val Thr Ile Val Asp Leu Leu Val Thr Val Leu Asn Leu
785 790 795 800
Leu Ala Ile Ser Leu Gly Tyr Phe Gly Pro Lys Cys Tyr Met Ile Leu
805 810 815
Phe Tyr Pro Glu Arg Asn Thr Pro Ala Tyr Phe Asn Ser Met Ile Gln
820 825 830
Gly Tyr Thr Met Arg Arg Asp
835

<210> 7
<211> 852
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence; note =
synthetic construct

<400> 7
Met Leu Gly Pro Ala Val Leu Gly Leu Ser Leu Trp Ala Leu Leu His

```
1               5                   10                  15
Pro Gly Thr Gly Ala Pro Leu Cys Leu Ser Gln Gln Leu Arg Met Lys
            20                  25                  30
Gly Asp Tyr Val Leu Gly Gly Leu Phe Pro Leu Gly Glu Ala Glu Glu
        35                  40                  45
Ala Gly Leu Arg Ser Arg Thr Arg Pro Ser Ser Pro Val Cys Thr Arg
    50                  55                  60
Phe Ser Ser Asn Gly Leu Leu Trp Ala Leu Ala Met Lys Met Ala Val
65                  70                  75                  80
Glu Glu Ile Asn Asn Lys Ser Asp Leu Leu Pro Gly Leu Arg Leu Gly
                85                  90                  95
Tyr Asp Leu Phe Asp Thr Cys Ser Glu Pro Val Val Ala Met Lys Pro
            100                 105                 110
Ser Leu Met Phe Leu Ala Lys Ala Gly Ser Arg Asp Ile Ala Ala Tyr
        115                 120                 125
Cys Asn Tyr Thr Gln Tyr Gln Pro Arg Val Leu Ala Val Ile Gly Pro
    130                 135                 140
His Ser Ser Glu Leu Ala Met Val Thr Gly Lys Phe Phe Ser Phe Phe
145                 150                 155                 160
Leu Met Pro Gln Val Ser Tyr Gly Ala Ser Met Glu Leu Leu Ser Ala
                165                 170                 175
Arg Glu Thr Phe Pro Ser Phe Phe Arg Thr Val Pro Ser Asp Arg Val
            180                 185                 190
Gln Leu Thr Ala Ala Ala Glu Leu Leu Gln Glu Phe Gly Trp Asn Trp
        195                 200                 205
Val Ala Ala Leu Gly Ser Asp Asp Glu Tyr Gly Arg Gln Gly Leu Ser
    210                 215                 220
Ile Phe Ser Ala Leu Ala Ala Ala Arg Gly Ile Cys Ile Ala His Glu
225                 230                 235                 240
Gly Leu Val Pro Leu Pro Arg Ala Asp Asp Ser Arg Leu Gly Lys Val
                245                 250                 255
Gln Asp Val Leu His Gln Val Asn Gln Ser Ser Val Gln Val Val Leu
            260                 265                 270
Leu Phe Ala Ser Val His Ala Ala His Ala Leu Phe Asn Tyr Ser Ile
        275                 280                 285
Ser Ser Arg Leu Ser Pro Lys Val Trp Val Ala Ser Glu Ala Trp Leu
    290                 295                 300
Thr Ser Asp Leu Val Met Gly Leu Pro Gly Met Ala Gln Met Gly Thr
305                 310                 315                 320
Val Leu Gly Phe Leu Gln Arg Gly Ala Gln Leu His Glu Phe Pro Gln
                325                 330                 335
Tyr Val Lys Thr His Leu Ala Leu Ala Thr Asp Pro Ala Phe Cys Ser
            340                 345                 350
Ala Leu Gly Glu Arg Glu Gln Gly Leu Glu Glu Asp Val Val Gly Gln
        355                 360                 365
Arg Cys Pro Gln Cys Asp Cys Ile Thr Leu Gln Asn Val Ser Ala Gly
    370                 375                 380
Leu Asn His His Gln Thr Phe Ser Val Tyr Ala Ala Val Tyr Ser Val
385                 390                 395                 400
Ala Gln Ala Leu His Asn Thr Leu Gln Cys Asn Ala Ser Gly Cys Pro
                405                 410                 415
Ala Gln Asp Pro Val Lys Pro Trp Gln Leu Leu Glu Asn Met Tyr Asn
            420                 425                 430
Leu Thr Phe His Val Gly Gly Leu Pro Leu Arg Phe Asp Ser Ser Gly
        435                 440                 445
Asn Val Asp Met Glu Tyr Asp Leu Lys Leu Trp Val Trp Gln Gly Ser
    450                 455                 460
```

Val Pro Arg Leu His Asp Val Gly Arg Phe Asn Gly Ser Leu Arg Thr
465 470 475 480
Glu Arg Leu Lys Ile Arg Trp His Thr Ser Asp Asn Gln Lys Pro Val
485 490 495
Ser Arg Cys Ser Arg Gln Cys Gln Glu Gly Gln Val Arg Arg Val Lys
500 505 510
Gly Phe His Ser Cys Cys Tyr Asp Cys Val Asp Cys Glu Ala Gly Ser
515 520 525
Tyr Arg Gln Asn Pro Asp Asp Ile Ala Cys Thr Phe Cys Gly Gln Asp
530 535 540
Glu Trp Ser Pro Glu Arg Ser Thr Arg Cys Phe Arg Arg Arg Ser Arg
545 550 555 560
Phe Leu Ala Trp Gly Glu Pro Ala Val Leu Leu Leu Leu Leu Leu Leu
565 570 575
Ser Leu Ala Leu Gly Leu Val Leu Ala Ala Leu Gly Leu Phe Val His
580 585 590
His Arg Asp Ser Pro Leu Val Gln Ala Ser Gly Gly Pro Leu Ala Cys
595 600 605
Phe Gly Leu Val Cys Leu Gly Leu Val Cys Leu Ser Val Leu Leu Phe
610 615 620
Pro Gly Gln Pro Ser Pro Ala Arg Cys Leu Ala Gln Gln Pro Leu Ser
625 630 635 640
His Leu Pro Leu Thr Gly Cys Leu Ser Thr Leu Phe Leu Gln Ala Ala
645 650 655
Glu Ile Phe Val Glu Ser Glu Leu Pro Leu Ser Trp Ala Asp Arg Leu
660 665 670
Ser Gly Cys Leu Arg Gly Pro Trp Ala Trp Leu Val Val Leu Leu Ala
675 680 685
Met Leu Val Glu Val Ala Leu Cys Thr Trp Tyr Leu Val Ala Phe Pro
690 695 700
Pro Glu Val Val Thr Asp Trp His Met Leu Pro Thr Glu Ala Leu Val
705 710 715 720
His Cys Arg Thr Arg Ser Trp Val Ser Phe Gly Leu Ala His Ala Thr
725 730 735
Asn Ala Thr Leu Ala Phe Leu Cys Phe Leu Gly Thr Phe Leu Val Arg
740 745 750
Ser Gln Pro Gly Arg Tyr Asn Arg Ala Arg Gly Leu Thr Phe Ala Met
755 760 765
Leu Ala Tyr Phe Ile Thr Trp Val Ser Phe Val Pro Leu Leu Ala Asn
770 775 780
Val Gln Val Val Leu Arg Pro Ala Val Gln Met Gly Ala Leu Leu Leu
785 790 795 800
Cys Val Leu Gly Ile Leu Ala Ala Phe His Leu Pro Arg Cys Tyr Leu
805 810 815
Leu Met Arg Gln Pro Gly Leu Asn Thr Pro Glu Phe Phe Leu Gly Gly
820 825 830
Gly Pro Gly Asp Ala Gln Gly Gln Asn Asp Gly Asn Thr Gly Asn Gln
835 840 845
Gly Lys His Glu
850

<210> 8
<211> 2526
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence; note = synthetic construct

<400> 8
atgctgctct gcacggctcg cctggtcggc ctgcagcttc tcatttcctg ctgctgggcc 60
tttgcctgcc atagcacgga gtcttctcct gacttcaccc tccccggaga ttacctcctg 120
gcaggcctgt tccctctcca ttctggctgt ctgcaggtga ggcacagacc cgaggtgacc 180
ctgtgtgaca ggtcttgtag cttcaatgag catggctacc acctcttcca ggctatgcgg 240
cttggggttg aggagataaa caactccacg gccctgctgc ccaacatcac cctggggtac 300
cagctgtatg atgtgtgttc tgactctgcc aatgtgtatg ccacgctgag agtgctctcc 360
ctgccagggc aacaccacat agagctccaa ggagaccttc tccactattc ccctacggtg 420
ctggcagtga ttgggcctga cagcaccaac cgtgctgcca ccacagccgc cctgctgagc 480
cctttcctgg tgcccatgat tagctatgcg gccagcagcg agacgctcag cgtgaagcgg 540
cagtatccct ctttcctgcg caccatcccc aatgacaagt accaggtgga gaccatggtg 600
ctgctgctgc agaagttcgg gtggacctgg atctctctgg ttggcagcag tgacgactat 660
gggcagctag ggtgcaggc actggagaac caggccactg gtcaggggat ctgcattgct 720
ttcaaggaca tcatgccctt ctctgcccag gtgggcgatg agaggatgca gtgcctcatg 780
cgccacctgg cccaggccgg ggccaccgtc gtggttgttt tttccagccg gcagttggcc 840
agggtgtttt tcgagtccgt ggtgctgacc aacctgactg gcaaggtgtg ggtcgcctca 900
gaagcctggg ccctctccag gcacatcact ggggtgcccg ggatccagcg cattgggatg 960
gtgctgggcg tggccatcca gaagagggct gtccctggcc tgaaggcgtt tgaagaagcc 1020
tatgcccggg cagacaagaa ggcccctagg ccttgccaca agggctcctg gtgcagcagc 1080
aatcagctct gcagagaatg ccaagctttc atggcacaca cgatgcccaa gctcaaagcc 1140
ttctccatga gttctgccta caacgcatac cgggctgtgt atgcggtggc ccatggcctc 1200
caccagctcc tgggctgtgc ctctggagct tgttccaggg gccgagtcta cccctggcag 1260
cttttggagc agatccacaa ggtgcatttc cttctacaca aggacactgt ggcgtttaat 1320
gacaacagag atcccctcag tagctataac ataattgcct gggactggaa tggacccaag 1380
tggaccttca cggtcctcgg ttcctccaca tggtctccag ttcagctaaa cataaatgag 1440
accaaaatcc agtggcacgg aaaggacaac caggtgccta agtctgtgtg ttccagcgac 1500
tgtcttgaag ggcaccagcg agtggttacg ggtttccatc actgctgctt tgagtgtgtg 1560
ccctgtgggg ctgggacctt cctcaacaag agtgacctct acagatgcca gccttgtggg 1620
aaagaagagt gggcacctga gggaagccag acctgcttcc cgcgcactgt ggtgtttttg 1680
gctttgcgtg agcacacctc ttgggtgctg ctggcagcta acacgctgct gctgctgctg 1740
ctgcttggga ctgctggcct gtttgcctgg cacctagaca cccctgtggt gaggtcagca 1800
gggggccgcc tgtgctttct tatgctgggc tccctggcag caggtagtgg cagcctctat 1860
ggcttctttg gggaacccac aaggcctgcg tgcttgctac gccaggccct ctttgccctt 1920
ggtttcacca tcttcctgtc ctgcctgaca gttcgctcat tccaactaat catcatcttc 1980
aagttttcca ccaaggtacc tacattctac cacgcctggg tccaaaacca cggtgctggc 2040
ctgtttgtga tgatcagctc agcggcccag ctgcttatct gtctaacttg gctggtggtg 2100
tggaccccac tgcctgctag ggaataccag cgcttccccc atctggtgat gcttgagtgc 2160
acagagacca actccctggg cttcatactg gccttcctct acaatggcct cctctccatc 2220
agtgcctttg cctgcagcta cctgggtaag gacttgccag agaactacaa cgaggccaaa 2280
tgtgtcacct tcagcctgct cttcaacttc gtgtcctgga tcgccttctt caccacggcc 2340
agcgtctacg acggcaagta cctgcctgcg gccaacatga tggctgggct gagcagcctg 2400
agcagcggct tcggtgggta ttttctgcct aagtgctacg tgatcctctg ccgcccagac 2460
ctcaacagca cagagcactt ccaggcctcc attcaggact acacgaggcg ctgcggctcc 2520
acctga 2526

<210> 9
<211> 2559
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence; note = synthetic construct

<400> 9
atgctgggcc ctgctgtcct gggcctcagc ctctgggctc tcctgcaccc tgggacgggg 60
gccccattgt gcctgtcaca gcaacttagg atgaaggggg actacgtgct gggggggctg 120
ttccccctgg gcgaggccga ggaggctggc ctccgcagcc ggacacggcc cagcagccct 180
gtgtgcacca ggttctcctc aaacggcctg ctctgggcac tggccatgaa aatggccgtg 240
gaggagatca acaacaagtc ggatctgctg cccgggctgc gcctgggcta cgacctcttt 300
gatacgtgct cggagcctgt ggtggccatg aagcccagcc tcatgttcct ggccaaggca 360
ggcagccgcg acatcgccgc ctactgcaac tacacgcagt accagccccg tgtgctggct 420
gtcatcgggc ccactcgtc agagctcgcc atggtcaccg gcaagttctt cagcttcttc 480
ctcatgcccc aggtcagcta cggtgctagc atggagctgc tgagcgcccg ggagaccttc 540
ccctccttct tccgcaccgt gcccagcgac cgtgtgcagc tgacggccgc cgcggagctg 600
ctgcaggagt tcggctggaa ctgggtggcc gccctgggca gcgacgacga gtacggccgg 660
cagggcctga gcatcttctc ggccctggcc gcggcacgcg gcatctgcat cgcgcacgag 720
ggcctggtgc cgctgccccg tgccgatgac tcgcggctgg ggaaggtgca ggacgtcctg 780
caccaggtga accagagcag cgtgcaggtg gtgctgctgt tcgcctccgt gcacgccgcc 840
cacgccctct tcaactacag catcagcagc aggctctcgc ccaaggtgtg ggtggccagc 900
gaggcctggc tgacctctga cctggtcatg gggctgcccg gcatggccca gatgggcacg 960
gtgcttggct tcctccagag ggtgcccag ctgcacgagt tcccccagta cgtgaagacg 1020
cacctggccc tggccaccga cccggccttc tgctctgccc tgggcgagag ggagcagggt 1080
ctggaggagg acgtggtggg ccagcgctgc ccgcagtgtg actgcatcac gctgcagaac 1140
gtgagcgcag ggctaaatca ccaccagacg ttctctgtct acgcagctgt gtatagcgtg 1200
gcccaggccc tgcacaacac tcttcagtgc aacgcctcag gctgccccgc gcaggacccc 1260
gtgaagccct ggcagctcct ggagaacatg tacaacctga ccttccacgt gggcgggctg 1320
ccgctgcggt tcgacagcag cggaaacgtg gacatggagt acgacctgaa gctgtgggtg 1380
tggcagggct cagtgcccag gctccacgac gtgggcaggt tcaacggcag cctcaggaca 1440
gagcgcctga agatccgctg gcacacgtct gacaaccaga agcccgtgtc ccggtgctcg 1500
cggcagtgcc aggagggcca ggtgcgccgg gtcaaggggt tccactcctg ctgctacgac 1560
tgtgtggact gcgaggcggg cagctaccgg caaaacccag acgacatcgc ctgcaccttt 1620
tgtggccagg atgagtggtc cccggagcga agcacacgct gcttccgccg caggtctcgg 1680
ttcctggcat ggggcgagcc ggctgtgctg ctgctgctcc tgctgctgag cctggcgctg 1740
ggccttgtgc tggctgcttt ggggctgttc gttcaccatc gggacagccc actggttcag 1800
gcctcggggg ggcccctggc ctgctttggc ctggtgtgcc tgggcctggt ctgcctcagc 1860
gtcctcctgt tccctggcca gcccagccct gcccgatgcc tggcccagca gcccttgtcc 1920
cacctcccgc tcacgggctg cctgagcaca ctcttcctgc aggcggccga gatcttcgtg 1980
gagtcagaac tgcctctgag ctgggcagac cggctgagtg gctgcctgcg gggggccctgg 2040
gcctggctgg tggtgctgct ggccatgctg gtggaggtcg cactgtgcac ctggtacctg 2100
gtggccttcc cgccggaggt ggtgacggac tggcacatgc tgcccacgga ggcgctggtg 2160
cactgccgca cacgctcctg ggtcagcttc ggcctagcgc acgccaccaa tgccacgctg 2220
gcctttctct gcttcctggg cactttcctg gtgcggagcc agccgggctg ctacaaccgt 2280
gcccgtggcc tcacctttgc catgctggcc tacttcatca cctgggtctc ctttgtgccc 2340
ctcctggcca atgtgcaggt ggtcctcagg cccgccgtgc agatggcgc cctcctgctc 2400
tgtgtcctgg gcatcctggc tgccttccac ctgcccaggt gttacctgct catgcggcag 2460
ccagggctca acacccccga gttcttcctg ggagggggcc ctggggatgc ccaaggccag 2520
aatgacggga acacaggaaa tcaggggaaa catgagtga 2559

<210> 10
<211> 2518
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence; note =
synthetic construct

<400> 10
atggggccca gggcaaagac catctgctcc ctgttcttcc tcctatgggt cctggctgag 60
ccggctgaga actcggactt ctacctgcct ggggattacc tcctgggtgg cctcttctcc 120

```
ctccatgcca acatgaaggg cattgttcac cttaacttcc tgcaggtgcc catgtgcaag    180
gagtatgaag tgaaggtgat aggctacaac ctcatgcagg ccatgcgctt cgcggtggag    240
gagatcaaca atgacagcag cctgctgcct ggtgtgctgc tgggctatga gatcgtggat    300
gtgtgctaca tctccaacaa tgtccagccg gtgctctact tcctggcaca cgaggacaac    360
ctccttccca tccaagagga ctacagtaac tacatttccc gtgtggtggc tgtcattggc    420
cctgacaact ccgagtctgt catgactgtg gccaattcct ctccctattt ctccttccac    480
agatcaccta cagcgccatc agcgatgagc tgcgagacaa ggtgcgcttc ccggctttgc    540
tgcgtaccac acccagcgcc gaccaccacg tcgaggccat ggtgcagctg atgctgcact    600
tccgctggaa ctggatcatt gtgctggtga gcagcgacac ctatggccgc gacaatggca    660
gctgcttggc gagcgcgtgg cccggcgcga catctgcatc gccttccagg agacgctgcc    720
cacactgcag cccaaccaga acatgacgtc agaggagcgc cagcgcctgg tgaccattgt    780
ggacaagctg cagcagagca cagcgcgcgt cgtggtcgtg ttctcgcccg acctgaccct    840
gtaccacttc ttcaatgagg tgctgcgcca gaacttcacg ggcgccgtgt ggatcgcctc    900
cgagtcctgg gccatcgacc cggtcctgca caacctcacg gagctgggcc acttgggcac    960
cttcctgggc atcaccatcc agagcgtgcc catcccgggc ttcagtgagt tccgcgagtg   1020
ggcccacag gctgggccgc cacccctcag caggaccagc cagagctata cctgcaacca   1080
ggagtgcgac aactgcctga acgccacctt gtccttcaac accattctca ggctctctgg   1140
ggagcgtgtc gtctacagcg tgtactctgc ggtctatgct gtggcccatg ccctgcacag   1200
cctcctcggc tgtgacaaaa gcacctgcac caagagggtg gtctacccct ggcagctgct   1260
tgaggagatc tggaaggtca acttcactct cctggaccac caaatcttct tcgacccgca   1320
aggggacgtg gctctgcact tggagattgt ccagtggcaa tgggaccgga gccagaatcc   1380
cttccagagc gtcgcctcct actacccct gcagcgacag ctgaagaaca tccaagacat    1440
ctcctggcac accgtcaaca acacgatccc tatgtccatg tgttccaaga ggtgccagtc   1500
agggcaaaag aagaagcctg tgggcatcca cgtctgctgc ttcgagtgca tcgactgcct   1560
tcccggcacc ttcctcaacc acactgaaga tgaatatgaa tgccaggcct gcccgaataa   1620
cgagtggtcc taccagagtg agacctcctg cttcaagcgg cagctggtct tcctggaatg   1680
gcatgaggca cccaccatcg ctgtggccct gctggccgcc ctgggcttcc tcagcaccct   1740
ggccatcctg gtgatattct ggaggcactt ccagacaccc atagttcgct cggctggggg   1800
ccccatgtgc ttcctgatgc tgacactgct gctggtggca tacatggtgg tcccggtgta   1860
cgtggggccg cccaaggtct ccacctgcct ctgccgccag gccctctttc cctctgctt    1920
cacaatttgc atctcctgta tcgccgtgcg ttctttccag atcgtctgcg ccttcaagat   1980
ggccagccgc ttcccacgcg cctacagcta ctgggtccgc taccaggggc cctacgtctc   2040
tatggcattt atcacggtac tcaaaatggt cattgtggta attggcatgc tggccacggg   2100
cctcagtccc accacccgta ctgaccccga tgaccccaag atcacaattg tctcctgtaa   2160
ccccaactac cgcaacagcc tgctgttcaa caccagcctg gacctgctgc tctcagtggt   2220
gggtttcagc ttcgcctaca tgggcaaaga gctgcccacc aactacaacg aggccaagtt   2280
catcaccctc agcatgacct tctatttcac ctcatccgtc tccctctgca ccttcatgtc   2340
tgcctacagc ggggtgctgg tcaccatcgt ggacctcttg gtcactgtgc tcaacctcct   2400
ggccatcagc ctgggctact tcggccccaa gtgctacatg atcctcttct acccggagcg   2460
caacacgccc gcctacttca acagcatgat ccagggctac accatgagga gggactag     2518
```

```
<210> 11
<211> 2577
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence; note =
      synthetic construct

<400> 11
atgccgggtt tggctatctt gggcctcagt ctggctgctt tcctggagct tgggatgggg     60
tcctctttgt gtctgtcaca gcaattcaag gcacaagggg actatatatt gggtggacta    120
tttcccctgg gcacaactga ggaggccact ctcaaccaga gaacacagcc caacggcatc    180
ctatgtacca ggttctcgcc ccttggtttg ttcctggcca tggctatgaa gatggctgta    240
gaggagatca acaatggatc tgccttgctc cctgggctgc gactgggcta tgacctgttt    300
gacacatgct cagagccagt ggtcaccatg aagcccagcc tcatgttcat ggccaaggtg    360
```

```
ggaagtcaaa gcattgctgc ctactgcaac tacacacagt accaaccccg tgtgctggct     420
gtcattggtc cccactcatc agagcttgcc ctcattacag gcaagttctt cagcttcttc     480
ctcatgccac aggtcagcta tagtgccagc atggatcggc taagtgaccg ggaaacattt     540
ccatccttct tccgcacagt gcccagtgac cgggtgcagc tgcaggccgt tgtgacactg     600
ttgcagaatt tcagctggaa ctgggtggct gccttaggta gtgatgatga ctatggccgg     660
gaaggtctga gcatctttc tggtctggcc aactcacgag gtatctgcat tgcacacgag     720
ggcctggtgc cacaacatga cactagtggc caacaattgg gcaaggtggt ggatgtgcta     780
cgccaagtga accaaagcaa agtacaggtg gtggtgctgt ttgcatctgc ccgtgctgtc     840
tactcccttt ttagctacag catccttcat gacctctcac ccaaggtatg ggtggccagt     900
gagtcctggc tgacctctga cctggtcatg acacttccca atattgcccg tgtgggcact     960
gttcttgggt ttctgcagcg cggtgcccta ctgcctgaat ttcccatta tgtggagact    1020
cgccttgccc tagctgctga cccaacattc tgtgcctccc tgaaagctga gttggatctg    1080
gaggagcgcg tgatggggcc acgctgttca caatgtgact acatcatgct acagaacctg    1140
tcatctgggc tgatgcagaa cctatcagct gggcagttgc accaccaaat atttgcaacc    1200
tatgcagctg tgtacagtgt ggctcaggcc cttcacaaca ccctgcagtg caatgtctca    1260
cattgccaca catcagagcc tgttcaaccc tggcagctcc tggagaacat gtacaatatg    1320
agtttccgtg ctcgagactt gacactgcag tttgatgcca aggggagtgt agacatggaa    1380
tatgacctga agatgtgggt gtggcagagc cctacacctg tactacatac tgtaggcacc    1440
ttcaacggca cccttcagct gcagcactcg aaaatgtatt ggccaggcaa ccaggtgcca    1500
gtctcccagt gctcccggca gtgcaaagat ggccaggtgc gcagagtaaa gggctttcat    1560
tcctgctgct atgactgtgt ggactgcaag gcagggagct accggaagca tccagatgac    1620
ttcacctgta ctccatgtgg caaggatcag tggtccccag aaaaaagcac aacctgctta    1680
cctcgcaggc ccaagtttct ggcttggggg gagccagctg tgctgtcact tctcctgctg    1740
ctttgcctgg tgctgggcct gacactggct gccctggggc tctttgtcca ctactgggac    1800
agccctcttg ttcaggcctc aggtgggtca ctgttctgct ttggcctgat ctgcctaggc    1860
ctcttctgcc tcagtgtcct tctgttccca ggacgaccac gctctgccag ctgccttgcc    1920
caacaaccaa tggctcacct ccctctcaca ggctgcctga gcacactctt cctgcaagca    1980
gccgagatct ttgtggagtc tgagctgcca ctgagttggg caaactggct ctgcagctac    2040
cttcggggcc cctgggcttg gctggtggta ctgctggcca ctcttgtgga ggctgcacta    2100
tgtgcctggt acttgatggc tttccctcca gaggtggtga cagattggca ggtgctgccc    2160
acggaggtac tggaacactg ccgcatgcgt tcctgggtca gcctgggctt ggtgcacatc    2220
accaatgcag tgttagcttt cctctgcttt ctgggcactt tcctggtaca gagccagcct    2280
ggtcgctata accgtgcccg tggcctcacc ttcgccatgc tagcttattt catcatctgg    2340
gtctcttttg tgcccctcct ggctaatgtg caggtggcct accagccagc tgtgcagatg    2400
ggtgctatct tattctgtgc cctgggcatc ctggccacct tccacctgcc caaatgctat    2460
gtacttctgt ggctgccaga gctcaacacc caggagttct tcctgggaag gagccccaag    2520
gaagcatcag atgggaatag tggtagtagt gaggcaactc ggggacacag tgaatga       2577
```

<210> 12
<211> 137
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence; note =
synthetic construct

<400> 12

```
Pro Ser Pro Phe Arg Asp Ile Val Ser Tyr Pro Asp Lys Ile Ile Leu
1               5                   10                  15
Gly Cys Phe Met Asn Leu Lys Thr Ser Ser Val Ser Phe Val Leu Leu
            20                  25                  30
Leu Leu Leu Cys Leu Leu Cys Phe Ile Phe Ser Tyr Met Gly Lys Asp
        35                  40                  45
Leu Pro Lys Asn Tyr Asn Glu Ala Lys Ala Ile Thr Phe Cys Leu Leu
    50                  55                  60
Leu Leu Ile Leu Thr Trp Ile Ile Phe Thr Thr Ala Ser Leu Leu Tyr
```

```
65                  70                  75                  80
Gln Gly Lys Tyr Ile His Ser Leu Asn Ala Leu Ala Val Leu Ser Ser
                85                  90                  95
Ile Tyr Ser Phe Leu Leu Trp Tyr Phe Leu Pro Lys Cys Tyr Ile Ile
            100                 105                 110
Ile Phe Gln Pro Gln Lys Asn Thr Gln Lys Tyr Phe Gln Gly Leu Ile
        115                 120                 125
Gln Asp Tyr Thr Lys Thr Ile Ser Gln
    130                 135
```

<210> 13
<211> 242
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence; note = synthetic construct

<400> 13

```
Phe Ala Val Asn Tyr Asn Thr Pro Val Val Arg Ser Ala Gly Gly Pro
 1              5                   10                  15
Met Cys Phe Leu Ile Leu Gly Cys Leu Ser Leu Cys Ser Ile Ser Val
            20                  25                  30
Phe Phe Tyr Phe Glu Arg Pro Thr Glu Ala Phe Cys Ile Leu Arg Phe
        35                  40                  45
Met Pro Phe Leu Leu Phe Tyr Ala Val Cys Leu Ala Cys Phe Ala Val
    50                  55                  60
Arg Ser Phe Gln Ile Val Ile Ile Phe Lys Ile Ala Ala Lys Phe Pro
65                  70                  75                  80
Arg Val His Ser Trp Trp Met Lys Tyr His Gly Gln Trp Leu Val Ile
                85                  90                  95
Ser Met Thr Phe Val Leu Gln Ala Val Val Ile Val Ile Gly Phe Ser
            100                 105                 110
Ser Asn Pro Pro Leu Pro Tyr Xaa Xaa Phe Val Ser Tyr Pro Asp Lys
        115                 120                 125
Ile Ile Leu Gly Cys Asp Val Asn Leu Asn Met Ala Ser Thr Ser Phe
    130                 135                 140
Phe Leu Leu Leu Leu Leu Cys Ile Leu Cys Phe Thr Phe Ser Tyr Met
145                 150                 155                 160
Gly Lys Asp Leu Pro Lys Asn Tyr Asn Glu Ala Lys Ala Ile Thr Phe
                165                 170                 175
Cys Leu Leu Leu Leu Ile Leu Thr Trp Ile Ile Phe Ala Thr Ala Phe
            180                 185                 190
Met Leu Tyr His Gly Lys Tyr Ile His Thr Leu Asn Ala Leu Ala Val
        195                 200                 205
Leu Ser Ser Ala Tyr Cys Phe Leu Leu Trp Tyr Phe Leu Pro Lys Cys
    210                 215                 220
Tyr Ile Ile Ile Phe Gln Pro His Lys Asn Thr Gln Lys Tyr Phe Gln

225                 230                 235                 240
Leu Ser
```

<210> 14
<211> 165
<212> PRT

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence; note = synthetic construct

<400> 14
```
Lys Lys Gln Gly Pro Glu Val Asp Ile Phe Ile Val Ser Val Thr Ile
 1               5                  10                  15
Leu Cys Ile Ser Val Leu Gly Val Ala Val Gly Pro Pro Glu Pro Ser
            20                  25                  30
Gln Asp Leu Asp Phe Tyr Met Asp Ser Ile Val Leu Glu Cys Ser Asn
        35                  40                  45
Thr Leu Ser Pro Gly Ser Phe Ile Glu Leu Cys Tyr Val Cys Val Leu
    50                  55                  60
Ser Val Leu Cys Phe Phe Phe Ser Tyr Met Gly Lys Asp Leu Pro Ala
65                  70                  75                  80
Asn Tyr Asn Glu Ala Lys Cys Val Thr Phe Ser Leu Met Val Tyr Met
                85                  90                  95
Ile Ser Trp Ile Ser Phe Phe Thr Val Tyr Leu Ile Ser Arg Gly Pro
            100                 105                 110
Phe Thr Val Ala Ala Tyr Val Cys Ala Thr Leu Val Ser Val Leu Ala
        115                 120                 125
Phe Phe Gly Gly Tyr Phe Leu Pro Lys Ile Tyr Ile Ile Val Leu Lys
    130                 135                 140
Pro Gln Met Asn Thr Thr Ala His Phe Gln Asn Cys Ile Gln Met Tyr
145                 150                 155                 160
Thr Met Ser Lys Gln
                165
```

<210> 15
<211> 236
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence; note = synthetic construct

<400> 15
```
Ala Pro Lys Ser Ser Gln Arg Xaa Leu Arg Arg Thr Arg Leu Xaa Leu
 1               5                  10                  15
Glu Trp Asp His Pro Met Ser Val Ala Leu Leu Phe Phe Leu Val Cys
            20                  25                  30
Cys Leu Leu Met Thr Ser Ser Ser Ala Val Ile Leu Leu Leu Asn Ile
        35                  40                  45
Asn Thr Pro Val Ala Lys Ser Ala Gly Gly Xaa Thr Cys Xaa Leu Lys
    50                  55                  60
Leu Ala Ala Leu Thr Ala Ala Ala Met Ser Ser Xaa Cys His Phe Gly
65                  70                  75                  80
Gln Pro Ser Pro Leu Ala Ser Lys Leu Lys Gln Pro Gln Phe Thr Phe
                85                  90                  95
Ser Phe Thr Val Cys Leu Ala Cys Asn Arg Cys Ala Leu Ala Thr Gly
            100                 105                 110
His Leu His Phe Xaa Ile Arg Val Ala Leu Pro Pro Ala Tyr Asn Xaa
        115                 120                 125
Trp Ala Lys Asn His Gly Pro Xaa Ala Thr Ile Phe Ile Ala Ser Ala
```

```
    130                 135                 140
Ala Ile Leu Cys Val Leu Cys Leu Arg Val Ala Val Gly Pro Pro Gln
145                 150                 155                 160
Pro Ser Gln Asx Leu Asx Phe Xaa Thr Asn Ser Ile Xaa Leu Xaa Xaa
                165                 170                 175
Ser Asn Thr Leu Ser Pro Gly Ser Phe Val Glu Leu Cys Asn Val Ser
            180                 185                 190
Leu Leu Ser Ala Val Cys Phe Val Phe Ser Xaa Met Gly Lys Asx Leu
        195                 200                 205
Pro Ala Asn Tyr Asn Glu Ala Lys Cys Val Thr Phe Ser Leu Met Val
    210                 215                 220
Asn Xaa Ile Ser Trp Ile Ser Phe Phe Thr Val Tyr
225                 230                 235
```

<210> 16
<211> 838
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence; note =
synthetic construct

<400> 16

```
Met Gly Pro Arg Ala Lys Thr Ile Cys Ser Leu Phe Phe Leu Leu Trp
 1               5                  10                  15
Val Leu Ala Glu Pro Ala Glu Asn Ser Asp Phe Tyr Leu Pro Gly Asp
            20                  25                  30
Tyr Leu Leu Gly Gly Leu Phe Ser Leu His Ala Asn Met Lys Gly Ile
        35                  40                  45
Val His Leu Asn Phe Leu Gln Val Pro Met Cys Lys Glu Tyr Glu Val
    50                  55                  60
Lys Val Ile Gly Tyr Asn Leu Met Gln Ala Met Arg Phe Ala Val Glu
65                  70                  75                  80
Glu Ile Asn Asn Asp Ser Ser Leu Leu Pro Gly Val Leu Leu Gly Tyr
                85                  90                  95
Glu Ile Val Asp Val Cys Tyr Ile Ser Asn Asn Val Gln Pro Val Leu
            100                 105                 110
Tyr Phe Leu Ala His Glu Asp Asn Leu Leu Pro Ile Gln Glu Asp Tyr
        115                 120                 125
Ser Asn Tyr Ile Ser Arg Val Val Ala Val Ile Gly Pro Asp Asn Ser
    130                 135                 140
Glu Ser Val Met Thr Val Ala Asn Phe Leu Ser Leu Phe Leu Leu Pro
145                 150                 155                 160
Gln Ile Thr Tyr Ser Ala Ile Ser Asp Glu Leu Arg Asp Lys Val Arg
                165                 170                 175
Phe Pro Ala Leu Leu Arg Thr Thr Pro Ser Ala Asp His His Val Glu
            180                 185                 190
Ala Met Val Gln Leu Met Leu His Phe Arg Trp Asn Trp Ile Ile Val
        195                 200                 205
Leu Val Ser Ser Asp Thr Tyr Gly Arg Asp Asn Gly Gln Leu Leu Gly
    210                 215                 220
Glu Arg Val Ala Arg Arg Asp Ile Cys Ile Ala Phe Gln Glu Thr Leu
225                 230                 235                 240
Pro Thr Leu Gln Pro Asn Gln Asn Met Thr Ser Glu Glu Arg Gln Arg
                245                 250                 255
Leu Val Thr Ile Val Asp Lys Leu Gln Gln Ser Thr Ala Arg Val Val
```

```
            260                 265                 270
Val Val Phe Ser Pro Asp Leu Thr Leu Tyr His Phe Phe Asn Glu Val
        275                 280                 285
Leu Arg Gln Asn Phe Thr Gly Ala Val Trp Ile Ala Ser Glu Ser Trp
    290                 295                 300
Ala Ile Asp Pro Val Leu His Asn Leu Thr Glu Leu Gly His Leu Gly
305                 310                 315                 320
Thr Phe Leu Gly Ile Thr Ile Gln Ser Val Pro Ile Pro Gly Phe Ser
                325                 330                 335
Glu Phe Arg Glu Trp Gly Pro Gln Ala Gly Pro Pro Pro Leu Ser Arg
            340                 345                 350
Thr Ser Gln Ser Tyr Thr Cys Asn Gln Glu Cys Asp Asn Cys Leu Asn
        355                 360                 365

Ala Thr Leu Ser Phe Asn Thr Ile Leu Arg Leu Ser Gly Glu Arg Val
    370                 375                 380
Val Tyr Ser Val Tyr Ser Ala Val Tyr Ala Val Ala His Ala Leu His
385                 390                 395                 400
Ser Leu Leu Gly Cys Asp Lys Ser Thr Cys Thr Lys Arg Val Val Tyr
                405                 410                 415
Pro Trp Gln Leu Leu Glu Glu Ile Trp Lys Val Asn Phe Thr Leu Leu
            420                 425                 430
Asp His Gln Ile Phe Phe Asp Pro Gln Gly Asp Val Ala Leu His Leu
        435                 440                 445
Glu Ile Val Gln Trp Gln Trp Asp Arg Ser Gln Asn Pro Phe Gln Ser
    450                 455                 460
Val Ala Ser Tyr Tyr Pro Leu Gln Arg Gln Leu Lys Asn Ile Gln Asp
465                 470                 475                 480
Ile Ser Trp His Thr Val Asn Asn Thr Ile Pro Met Ser Met Cys Ser
                485                 490                 495
Lys Arg Cys Gln Ser Gly Gln Lys Lys Lys Pro Val Gly Ile His Val
            500                 505                 510
Cys Cys Phe Glu Cys Ile Asp Cys Leu Pro Gly Thr Phe Leu Asn His
        515                 520                 525
Thr Glu Asp Glu Tyr Glu Cys Gln Ala Cys Pro Asn Asn Glu Trp Ser
    530                 535                 540
Tyr Gln Ser Glu Thr Ser Cys Phe Lys Arg Gln Leu Val Phe Leu Glu
545                 550                 555                 560
His Glu Val Pro Thr Ile Val Val Ala Ile Leu Ala Ala Leu Gly Phe
                565                 570                 575
Phe Ser Thr Leu Ala Ile Leu Phe Ile Phe Trp Arg His Phe Gln Thr
            580                 585                 590
Pro Met Val Arg Ser Ala Gly Gly Pro Met Cys Phe Leu Met Leu Val
        595                 600                 605
Pro Leu Leu Leu Ala Phe Gly Met Val Pro Val Tyr Val Gly Pro Pro
    610                 615                 620
Thr Val Phe Ser Cys Phe Cys Arg Gln Ala Phe Phe Thr Val Cys Phe
625                 630                 635                 640
Ser Ile Cys Leu Ser Cys Ile Thr Val Arg Ser Phe Gln Ile Val Cys
                645                 650                 655
Val Phe Lys Met Ala Arg Arg Leu Pro Ser Ala Tyr Ser Phe Trp Met
            660                 665                 670
Arg Tyr His Gly Pro Tyr Val Phe Val Ala Phe Ile Thr Ala Ile Lys
        675                 680                 685
Val Ala Leu Val Val Gly Asn Met Leu Ala Thr Thr Ile Asn Pro Ile
    690                 695                 700
Gly Arg Thr Asp Pro Asp Asp Pro Asn Ile Met Ile Leu Ser Cys His
```

```
705                 710                 715                 720
Pro Asn Tyr Arg Asn Gly Leu Leu Phe Asn Thr Ser Met Asp Leu Leu
                725                 730                 735
Leu Ser Val Leu Gly Phe Ser Phe Ala Tyr Met Gly Lys Glu Leu Pro
            740                 745                 750
Thr Asn Tyr Asn Glu Ala Lys Phe Ile Thr Leu Ser Met Thr Phe Ser
        755                 760                 765
Phe Thr Ser Ser Ile Ser Leu Cys Thr Phe Met Ser Val His Asp Gly
    770                 775                 780
Val Leu Val Thr Ile Met Asp Leu Leu Val Thr Val Leu Asn Phe Leu
785                 790                 795                 800
Ala Ile Gly Leu Gly Tyr Phe Gly Pro Lys Cys Tyr Met Ile Leu Phe
                805                 810                 815
Tyr Pro Glu Arg Asn Thr Ser Ala Tyr Phe Asn Ser Met Ile Gln Gly
            820                 825                 830
Tyr Thr Met Arg Lys Ser
        835

&lt;210&gt; 17
&lt;211&gt; 844
&lt;212&gt; PRT
&lt;213&gt; Artificial Sequence

&lt;220&gt;
&lt;223&gt; Description of Artificial Sequence; note =
      synthetic construct

&lt;400&gt; 17
Met Gly Pro Gln Ala Arg Thr Leu Cys Leu Leu Ser Leu Leu Leu His
 1               5                  10                  15
Val Leu Pro Lys Pro Gly Lys Leu Val Glu Asn Ser Asp Phe His Leu
            20                  25                  30
Ala Gly Asp Tyr Leu Leu Gly Gly Leu Phe Thr Leu His Ala Asn Val
        35                  40                  45
Lys Ser Ile Ser His Leu Ser Tyr Leu Gln Val Pro Lys Cys Asn Glu
    50                  55                  60
Phe Thr Met Lys Val Leu Gly Tyr Asn Leu Met Gln Ala Met Arg Phe
65                  70                  75                  80
Ala Val Glu Glu Ile Asn Asn Cys Ser Ser Leu Leu Pro Gly Val Leu
                85                  90                  95
Leu Gly Tyr Glu Met Val Asp Val Cys Tyr Leu Ser Asn Asn Ile His
            100                 105                 110
Pro Gly Leu Tyr Phe Leu Ala Gln Asp Asp Asp Leu Leu Pro Ile Leu
        115                 120                 125
Lys Asp Tyr Ser Gln Tyr Met Pro His Val Val Ala Val Ile Gly Pro
    130                 135                 140
Asp Asn Ser Glu Ser Ala Ile Thr Val Ser Asn Ile Leu Ser His Phe
145                 150                 155                 160
Leu Ile Pro Gln Ile Thr Tyr Ser Ala Ile Ser Asp Lys Leu Arg Asp
                165                 170                 175
Lys Arg His Phe Pro Ser Met Leu Arg Thr Val Pro Ser Ala Thr His
            180                 185                 190
His Ile Glu Ala Met Val Gln Leu Met Val His Phe Gln Trp Asn Trp
        195                 200                 205
Ile Val Val Leu Val Ser Asp Asp Asp Tyr Gly Arg Glu Asn Ser His
    210                 215                 220
Leu Leu Ser Gln Arg Leu Thr Lys Thr Ser Asp Ile Cys Ile Ala Phe
```

```
225                 230                 235                 240
Gln Glu Val Leu Pro Ile Pro Glu Ser Ser Gln Val Met Arg Ser Glu
                245                 250                 255
Glu Gln Arg Gln Leu Asp Asn Ile Leu Asp Lys Leu Arg Arg Thr Ser
            260                 265                 270
Ala Arg Val Val Val Val Phe Ser Pro Glu Leu Ser Leu Tyr Ser Phe
        275                 280                 285
Phe His Glu Val Leu Arg Trp Asn Phe Thr Gly Phe Val Trp Ile Ala
    290                 295                 300
Ser Glu Ser Trp Ala Ile Asp Pro Val Leu His Asn Leu Thr Glu Leu
305                 310                 315                 320
Arg His Thr Gly Thr Phe Leu Gly Val Thr Ile Gln Arg Val Ser Ile
                325                 330                 335
Pro Gly Phe Ser Gln Phe Arg Val Arg Arg Asp Lys Pro Gly Tyr Pro
            340                 345                 350
Val Pro Asn Thr Thr Asn Leu Arg Thr Thr Cys Asn Gln Asp Cys Asp
        355                 360                 365
Ala Cys Leu Asn Thr Thr Lys Ser Phe Asn Asn Ile Leu Ile Leu Ser
    370                 375                 380
Gly Glu Arg Val Val Tyr Ser Val Tyr Ser Ala Val Tyr Ala Val Ala
385                 390                 395                 400
His Ala Leu His Arg Leu Leu Gly Cys Asn Arg Val Arg Cys Thr Lys
                405                 410                 415
Gln Lys Val Tyr Pro Trp Gln Leu Leu Arg Glu Ile Trp His Val Asn
            420                 425                 430
Phe Thr Leu Leu Gly Asn Arg Leu Phe Phe Asp Gln Gln Gly Asp Met
        435                 440                 445
Pro Met Leu Leu Asp Ile Ile Gln Trp Gln Trp Asp Leu Ser Gln Asn
    450                 455                 460
Pro Phe Gln Ser Ile Ala Ser Tyr Ser Pro Thr Ser Lys Arg Leu Thr
465                 470                 475                 480
Tyr Ile Asn Asn Val Ser Trp Tyr Thr Pro Asn Asn Thr Val Pro Val
                485                 490                 495
Ser Met Cys Ser Lys Ser Cys Gln Pro Gly Gln Met Lys Lys Ser Val
            500                 505                 510
Gly Leu His Pro Cys Cys Phe Glu Cys Leu Asp Cys Met Pro Gly Thr
        515                 520                 525
Tyr Leu Asn Arg Ser Ala Asp Glu Phe Asn Cys Leu Ser Cys Pro Gly
    530                 535                 540
Ser Met Trp Ser Tyr Lys Asn Asp Ile Thr Cys Phe Gln Arg Arg Pro
545                 550                 555                 560
Thr Phe Leu Glu Trp Trp His Glu Ala Pro Thr Ile Ala Val Ala Leu
                565                 570                 575
Leu Ala Ala Leu Gly Phe Leu Ser Thr Leu Ala Ile Leu Val Ile Phe
            580                 585                 590
Trp Arg His Phe Gln Thr Pro Ile Val Arg Ser Ala Gly Gly Pro Met
        595                 600                 605
Cys Phe Leu Met Leu Thr Leu Leu Leu Val Ala Tyr Met Val Val Pro
    610                 615                 620
Val Tyr Val Gly Pro Pro Lys Val Ser Thr Cys Leu Cys Arg Gln Ala
625                 630                 635                 640
Leu Phe Pro Leu Cys Phe Thr Ile Cys Ile Ser Cys Ile Ala Val Arg
                645                 650                 655
Ser Phe Gln Ile Val Cys Ala Phe Lys Met Ala Ser Arg Phe Pro Arg
            660                 665                 670
Ala Tyr Ser Tyr Trp Val Arg Tyr Gln Gly Pro Tyr Val Ser Met Ala
        675                 680                 685
```

```
Phe Ile Thr Val Leu Lys Met Val Ile Val Val Ile Gly Met Leu Ala
    690                 695                 700
Thr Gly Leu Ser Pro Thr Thr Arg Thr Asp Pro Asp Asp Pro Lys Ile
705                 710                 715                 720
Thr Ile Val Ser Cys Asn Pro Asn Tyr Arg Asn Ser Leu Leu Phe Asn
                725                 730                 735
Thr Ser Leu Asp Leu Leu Leu Ser Val Val Gly Phe Ser Phe Ala Tyr
            740                 745                 750

Met Gly Lys Glu Leu Pro Thr Asn Tyr Asn Glu Ala Lys Phe Ile Thr
        755                 760                 765
Leu Ser Met Thr Phe Tyr Phe Thr Ser Ser Val Ser Leu Cys Thr Phe
    770                 775                 780
Met Ser Ala Tyr Ser Gly Val Leu Val Thr Ile Val Asp Leu Leu Val
785                 790                 795                 800
Thr Val Leu Asn Leu Leu Ala Ile Ser Leu Gly Tyr Phe Gly Pro Lys
                805                 810                 815
Cys Tyr Met Ile Leu Phe Tyr Pro Glu Arg Asn Thr Pro Ala Tyr Phe
            820                 825                 830
Asn Ser Met Ile Gln Gly Tyr Thr Met Arg Arg Asp
        835                 840
```

<210> 18
<211> 855
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence; note = synthetic construct

<400> 18

```
Met Leu Gly Pro Ala Val Leu Gly Leu Ser Leu Trp Ala Leu Leu His
 1               5                  10                  15
Pro Gly Thr Gly Ala Pro Leu Cys Leu Ser Gln Gln Leu Arg Met Lys
            20                  25                  30
Gly Asp Tyr Val Leu Gly Gly Leu Phe Pro Leu Gly Glu Ala Glu Glu
        35                  40                  45
Ala Gly Leu Arg Ser Arg Thr Arg Pro Ser Ser Pro Val Cys Thr Arg

    50                  55                  60
Phe Ser Ser Asn Gly Leu Leu Trp Ala Leu Ala Met Lys Met Ala Val
65                  70                  75                  80
Glu Glu Ile Asn Asn Lys Ser Asp Leu Leu Pro Gly Leu Arg Leu Gly
                85                  90                  95
Tyr Asp Leu Phe Asp Thr Cys Ser Glu Pro Val Val Ala Met Lys Pro
            100                 105                 110
Ser Leu Met Phe Leu Ala Lys Ala Gly Ser Arg Asp Ile Ala Ala Tyr
        115                 120                 125
Cys Asn Tyr Thr Gln Tyr Gln Pro Arg Val Leu Ala Val Ile Gly Pro
    130                 135                 140
His Ser Ser Glu Leu Ala Met Val Thr Gly Lys Phe Phe Ser Phe Phe
145                 150                 155                 160
Leu Met Pro Gln Val Ser Tyr Gly Ala Ser Met Glu Leu Leu Ser Ala
                165                 170                 175
Arg Glu Thr Phe Pro Ser Phe Phe Arg Thr Val Pro Ser Asp Arg Val
            180                 185                 190
```

Gln Leu Thr Ala Ala Ala Glu Leu Leu Gln Glu Phe Gly Trp Asn Trp
195 200 205
Val Ala Ala Leu Gly Ser Asp Asp Glu Tyr Gly Arg Gln Gly Leu Ser
210 215 220
Ile Phe Ser Ala Leu Ala Ala Ala Arg Gly Ile Cys Ile Ala His Glu
225 230 235 240
Gly Leu Val Pro Leu Pro Arg Ala Asp Asp Ser Arg Leu Gly Lys Val
245 250 255
Gln Asp Val Leu His Gln Val Asn Gln Ser Ser Val Gln Val Val Leu
260 265 270
Leu Phe Ala Ser Val His Ala Ala His Ala Leu Phe Asn Tyr Ser Ile
275 280 285
Ser Ser Arg Leu Ser Pro Lys Val Trp Val Ala Ser Glu Ala Trp Leu
290 295 300
Thr Ser Asp Leu Val Met Gly Leu Pro Gly Met Ala Gln Met Gly Thr
305 310 315 320
Val Leu Gly Phe Leu Gln Arg Gly Ala Gln Leu His Glu Phe Pro Gln
325 330 335
Tyr Val Lys Thr His Leu Ala Leu Ala Thr Asp Pro Ala Phe Cys Ser
340 345 350
Ala Leu Gly Glu Arg Glu Gln Gly Leu Glu Glu Asp Val Val Gly Gln
355 360 365
Arg Cys Pro Gln Cys Asp Cys Ile Thr Leu Gln Asn Val Ser Ala Gly
370 375 380
Leu Asn His His Gln Thr Phe Ser Val Tyr Ala Ala Val Tyr Ser Val
385 390 395 400
Ala Gln Ala Leu His Asn Thr Leu Gln Cys Asn Ala Ser Gly Cys Pro
405 410 415
Ala Gln Asp Pro Val Lys Pro Trp Gln Leu Leu Glu Asn Met Tyr Asn
420 425 430
Leu Thr Phe His Val Gly Gly Leu Pro Leu Arg Phe Asp Ser Ser Gly
435 440 445
Asn Val Asp Met Glu Tyr Asp Leu Lys Leu Trp Val Trp Gln Gly Ser
450 455 460
Val Pro Arg Leu His Asp Val Gly Arg Phe Asn Gly Ser Leu Arg Thr
465 470 475 480
Glu Arg Leu Lys Ile Arg Trp His Thr Ser Asp Asn Gln Lys Pro Val
485 490 495
Ser Arg Cys Ser Arg Gln Cys Gln Glu Gly Gln Val Arg Arg Val Lys
500 505 510
Gly Phe His Ser Cys Cys Tyr Asp Cys Val Asp Cys Glu Ala Gly Ser
515 520 525
Tyr Arg Gln Asn Pro Asp Asp Ile Ala Cys Thr Phe Cys Gly Gln Asp
530 535 540
Glu Trp Ser Pro Glu Arg Ser Thr Arg Cys Phe Arg Arg Arg Ser Arg
545 550 555 560
Phe Leu Glu Leu Ala Trp Gly Glu Pro Ala Val Leu Ser Leu Leu Leu
565 570 575
Leu Leu Cys Leu Val Leu Gly Leu Thr Leu Ala Ala Leu Gly Leu Phe
580 585 590
Val His Tyr Trp Asp Ser Pro Leu Val Gln Ala Ser Gly Gly Ser Leu
595 600 605
Phe Cys Phe Gly Leu Ile Cys Leu Gly Leu Phe Cys Leu Ser Val Leu
610 615 620
Leu Phe Pro Gly Arg Pro Arg Ser Ala Ser Cys Leu Ala Gln Gln Pro
625 630 635 640
Met Ala His Leu Pro Leu Thr Gly Cys Leu Ser Thr Leu Phe Leu Gln

```
                645                 650                 655
Ala Ala Glu Ile Phe Val Glu Ser Glu Leu Pro Leu Ser Trp Ala Asn
            660                 665                 670
Trp Leu Cys Ser Tyr Leu Arg Gly Pro Trp Ala Trp Leu Val Val Leu
        675                 680                 685
Leu Ala Thr Leu Val Glu Ala Ala Leu Cys Ala Trp Tyr Leu Met Ala
    690                 695                 700
Phe Pro Pro Glu Val Val Thr Asp Trp Gln Val Leu Pro Thr Glu Val
705                 710                 715                 720
Leu Glu His Cys Arg Met Arg Ser Trp Val Ser Leu Gly Leu Val His
                725                 730                 735
Ile Thr Asn Ala Val Leu Ala Phe Leu Cys Phe Leu Gly Thr Phe Leu
            740                 745                 750
Val Gln Ser Gln Pro Gly Arg Tyr Asn Arg Ala Arg Gly Leu Thr Phe
        755                 760                 765
Ala Met Leu Ala Tyr Phe Ile Ile Trp Val Ser Phe Val Pro Leu Leu
    770                 775                 780
Ala Asn Val Gln Val Ala Tyr Gln Pro Ala Val Gln Met Gly Ala Ile
785                 790                 795                 800
Leu Phe Cys Ala Leu Gly Ile Leu Ala Thr Phe His Leu Pro Lys Cys
                805                 810                 815
Tyr Val Leu Leu Trp Leu Pro Glu Leu Asn Thr Gln Glu Phe Phe Leu
            820                 825                 830
Gly Arg Ser Pro Lys Glu Ala Ser Asp Gly Asn Ser Gly Ser Ser Glu
        835                 840                 845
Ala Thr Arg Gly His Ser Glu
    850                 855

<210> 19
<211> 859
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence; note =
      synthetic construct

<400> 19
Met Pro Gly Leu Ala Ile Leu Gly Leu Ser Leu Ala Ala Phe Leu Glu
 1               5                  10                  15
Leu Gly Met Gly Ser Ser Leu Cys Leu Ser Gln Gln Phe Lys Ala Gln
            20                  25                  30
Gly Asp Tyr Ile Leu Gly Gly Leu Phe Pro Leu Gly Thr Thr Glu Glu
        35                  40                  45
Ala Thr Leu Asn Gln Arg Thr Gln Pro Asn Gly Ile Leu Cys Thr Arg
    50                  55                  60
Phe Ser Pro Leu Gly Leu Phe Leu Ala Met Ala Met Lys Met Ala Val
65                  70                  75                  80
Glu Glu Ile Asn Asn Gly Ser Ala Leu Leu Pro Gly Leu Arg Leu Gly
                85                  90                  95
Tyr Asp Leu Phe Asp Thr Cys Ser Glu Pro Val Val Thr Met Lys Pro
            100                 105                 110
Ser Leu Met Phe Met Ala Lys Val Gly Ser Gln Ser Ile Ala Ala Tyr
        115                 120                 125
Cys Asn Tyr Thr Gln Tyr Gln Pro Arg Val Leu Ala Val Ile Gly Pro
    130                 135                 140
His Ser Ser Glu Leu Ala Leu Ile Thr Gly Lys Phe Phe Ser Phe Phe
```

```
145                 150                 155                 160
Leu Met Pro Gln Val Ser Tyr Ser Ala Ser Met Asp Arg Leu Ser Asp
                165                 170                 175
Arg Glu Thr Phe Pro Ser Phe Phe Arg Thr Val Pro Ser Asp Arg Val
            180                 185                 190
Gln Leu Gln Ala Val Val Thr Leu Leu Gln Asn Phe Ser Trp Asn Trp
        195                 200                 205
Val Ala Ala Leu Gly Ser Asp Asp Asp Tyr Gly Arg Glu Gly Leu Ser
    210                 215                 220
Ile Phe Ser Gly Leu Ala Asn Ser Arg Gly Ile Cys Ile Ala His Glu
225                 230                 235                 240
Gly Leu Val Pro Gln His Asp Thr Ser Gly Gln Gln Leu Gly Lys Val
                245                 250                 255
Val Asp Val Leu Arg Gln Val Asn Gln Ser Lys Val Gln Val Val Val
            260                 265                 270
Leu Phe Ala Ser Ala Arg Ala Val Tyr Ser Leu Phe Ser Tyr Ser Ile
        275                 280                 285
Leu His Asp Leu Ser Pro Lys Val Trp Val Ala Ser Glu Ser Trp Leu
    290                 295                 300
Thr Ser Asp Leu Val Met Thr Leu Pro Asn Ile Ala Arg Val Gly Thr
305                 310                 315                 320
Val Leu Gly Phe Leu Gln Arg Gly Ala Leu Leu Pro Glu Phe Ser His
                325                 330                 335
Tyr Val Glu Thr Arg Leu Ala Leu Ala Ala Asp Pro Thr Phe Cys Ala
            340                 345                 350
Ser Leu Lys Ala Glu Leu Asp Leu Glu Glu Arg Val Met Gly Pro Arg
        355                 360                 365
Cys Ser Gln Cys Asp Tyr Ile Met Leu Gln Asn Leu Ser Ser Gly Leu
    370                 375                 380
Met Gln Asn Leu Ser Ala Gly Gln Leu His His Gln Ile Phe Ala Thr
385                 390                 395                 400
Tyr Ala Ala Val Tyr Ser Val Ala Gln Ala Leu His Asn Thr Leu Gln
                405                 410                 415
Cys Asn Val Ser His Cys His Thr Ser Glu Pro Val Gln Pro Trp Gln
            420                 425                 430
Leu Leu Glu Asn Met Tyr Asn Met Ser Phe Arg Ala Arg Asp Leu Thr
        435                 440                 445
Leu Gln Phe Asp Ala Lys Gly Ser Val Asp Met Glu Tyr Asp Leu Lys
    450                 455                 460
Met Trp Val Trp Gln Ser Pro Thr Pro Val Leu His Thr Val Gly Thr
465                 470                 475                 480
Phe Asn Gly Thr Leu Gln Leu Gln His Ser Lys Met Tyr Trp Pro Gly
                485                 490                 495
Asn Gln Val Pro Val Ser Gln Cys Ser Arg Gln Cys Lys Asp Gly Gln
            500                 505                 510
Val Arg Arg Val Lys Gly Phe His Ser Cys Cys Tyr Asp Cys Val Asp
        515                 520                 525
Cys Lys Ala Gly Ser Tyr Arg Lys His Pro Asp Asp Phe Thr Cys Thr
    530                 535                 540
Pro Cys Gly Lys Asp Gln Trp Ser Pro Glu Lys Ser Thr Thr Cys Leu
545                 550                 555                 560
Pro Arg Arg Pro Lys Phe Leu Glu Leu Ala Trp Gly Glu Pro Ala Val
                565                 570                 575
Leu Leu Leu Leu Leu Leu Leu Ser Leu Ala Leu Gly Leu Val Leu Ala
            580                 585                 590
Ala Leu Gly Leu Phe Val His His Arg Asp Ser Pro Leu Val Gln Ala
        595                 600                 605
```

Ser Gly Gly Pro Leu Ala Cys Phe Gly Leu Val Cys Leu Gly Leu Val
    610   615   620
Cys Leu Ser Val Leu Leu Phe Pro Gly Gln Pro Ser Pro Ala Arg Cys
625   630   635   640
Leu Ala Gln Gln Pro Leu Ser His Leu Pro Leu Thr Gly Cys Leu Ser
                645   650   655
Thr Leu Phe Leu Gln Ala Ala Glu Ile Phe Val Glu Ser Glu Leu Pro
            660   665   670
Leu Ser Trp Ala Asp Arg Leu Ser Gly Cys Leu Arg Gly Pro Trp Ala
        675   680   685
Trp Leu Val Val Leu Leu Ala Met Leu Val Glu Val Ala Leu Cys Thr
    690   695   700
Trp Tyr Leu Val Ala Phe Pro Pro Glu Val Val Thr Asp Trp His Met
705   710   715   720
Leu Pro Thr Glu Ala Leu Val His Cys Arg Thr Arg Ser Trp Val Ser
                725   730   735
Phe Gly Leu Ala His Ala Thr Asn Ala Thr Leu Ala Phe Leu Cys Phe
            740   745   750
Leu Gly Thr Phe Leu Val Arg Ser Gln Pro Gly Arg Tyr Asn Arg Ala
        755   760   765
Arg Gly Leu Thr Phe Ala Met Leu Ala Tyr Phe Ile Thr Trp Val Ser
    770   775   780
Phe Val Pro Leu Leu Ala Asn Val Gln Val Val Leu Arg Pro Ala Val
785   790   795   800
Gln Met Gly Ala Leu Leu Leu Cys Val Leu Gly Ile Leu Ala Ala Phe
                805   810   815
His Leu Pro Arg Cys Tyr Leu Leu Met Arg Gln Pro Gly Leu Asn Thr
            820   825   830
Pro Glu Phe Phe Leu Gly Gly Gly Pro Gly Asp Ala Gln Gly Gln Asn
        835   840   845
Asp Gly Asn Thr Gly Asn Gln Gly Lys His Glu
    850   855

<210> 20
<211> 841
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence; note =
      synthetic construct

<400> 20
Met Leu Leu Cys Thr Ala Arg Leu Val Gly Leu Gln Leu Leu Ile Ser
 1   5   10   15
Cys Cys Trp Ala Phe Ala Cys His Ser Thr Glu Ser Ser Pro Asp Phe
            20   25   30
Thr Leu Pro Gly Asp Tyr Leu Leu Ala Gly Leu Phe Pro Leu His Ser
        35   40   45
Gly Cys Leu Gln Val Arg His Arg Pro Glu Val Thr Leu Cys Asp Arg
    50   55   60
Ser Cys Ser Phe Asn Glu His Gly Tyr His Leu Phe Gln Ala Met Arg
65   70   75   80
Leu Gly Val Glu Glu Ile Asn Asn Ser Thr Ala Leu Leu Pro Asn Ile
                85   90   95
Thr Leu Gly Tyr Gln Leu Tyr Asp Val Cys Ser Asp Ser Ala Asn Val
            100   105   110

```
Tyr Ala Thr Leu Arg Val Leu Ser Leu Pro Gly Gln His His Ile Glu
        115                 120                 125
Leu Gln Gly Asp Leu Leu His Tyr Ser Pro Thr Val Leu Ala Val Ile
    130                 135                 140
Gly Pro Asp Ser Thr Asn Arg Ala Ala Thr Thr Ala Ala Leu Leu Ser
145                 150                 155                 160
Pro Phe Leu Val Pro Met Ile Ser Tyr Ala Ala Ser Ser Glu Thr Leu
                165                 170                 175
Ser Val Lys Arg Gln Tyr Pro Ser Phe Leu Arg Thr Ile Pro Asn Asp
            180                 185                 190
Lys Tyr Gln Val Glu Thr Met Val Leu Leu Leu Gln Lys Phe Gly Trp
        195                 200                 205
Thr Trp Ile Ser Leu Val Gly Ser Ser Asp Asp Tyr Gly Gln Leu Gly
    210                 215                 220
Val Gln Ala Leu Glu Asn Gln Ala Thr Gly Gln Gly Ile Cys Ile Ala
225                 230                 235                 240
Phe Lys Asp Ile Met Pro Phe Ser Ala Gln Val Gly Asp Glu Arg Met
                245                 250                 255
Gln Cys Leu Met Arg His Leu Ala Gln Ala Gly Ala Thr Val Val Val
            260                 265                 270
Val Phe Ser Ser Arg Gln Leu Ala Arg Val Phe Phe Glu Ser Val Val
        275                 280                 285
Leu Thr Asn Leu Thr Gly Lys Val Trp Val Ala Ser Glu Ala Trp Ala
    290                 295                 300
Leu Ser Arg His Ile Thr Gly Val Pro Gly Ile Gln Arg Ile Gly Met
305                 310                 315                 320
Val Leu Gly Val Ala Ile Gln Lys Arg Ala Val Pro Gly Leu Lys Ala
                325                 330                 335
Phe Glu Glu Ala Tyr Ala Arg Ala Asp Lys Lys Ala Pro Arg Pro Cys
            340                 345                 350
His Lys Gly Ser Trp Cys Ser Ser Asn Gln Leu Cys Arg Glu Cys Gln
        355                 360                 365
Ala Phe Met Ala His Thr Met Pro Lys Leu Lys Ala Phe Ser Met Ser
    370                 375                 380
Ser Ala Tyr Asn Ala Tyr Arg Ala Val Tyr Ala Val Ala His Gly Leu
385                 390                 395                 400
His Gln Leu Leu Gly Cys Ala Ser Gly Ala Cys Ser Arg Gly Arg Val
                405                 410                 415
Tyr Pro Trp Gln Leu Leu Glu Gln Ile His Lys Val His Phe Leu Leu
            420                 425                 430
His Lys Asp Thr Val Ala Phe Asn Asp Asn Arg Asp Pro Leu Ser Ser
        435                 440                 445
Tyr Asn Ile Ile Ala Trp Asp Trp Asn Gly Pro Lys Trp Thr Phe Thr
    450                 455                 460
Val Leu Gly Ser Ser Thr Trp Ser Pro Val Gln Leu Asn Ile Asn Glu
465                 470                 475                 480
Thr Lys Ile Gln Trp His Gly Lys Asp Asn Gln Val Pro Lys Ser Val
                485                 490                 495
Cys Ser Ser Asp Cys Leu Glu Gly His Gln Arg Val Val Thr Gly Phe
            500                 505                 510
His His Cys Cys Phe Glu Cys Val Pro Cys Gly Ala Gly Thr Phe Leu
        515                 520                 525
Asn Lys Ser Asp Leu Tyr Arg Cys Gln Pro Cys Gly Lys Glu Glu Trp
    530                 535                 540
Ala Pro Glu Gly Ser Gln Thr Cys Phe Pro Arg Thr Val Val Phe Leu
545                 550                 555                 560
Glu Trp His Glu Pro Ile Ser Leu Val Leu Ile Ala Ala Asn Thr Leu
```

```
                565                 570                 575
Leu Leu Leu Leu Leu Val Gly Thr Ala Gly Leu Phe Ala Trp His Phe
            580                 585                 590
His Thr Pro Val Val Arg Ser Ala Gly Gly Arg Leu Cys Phe Leu Met
        595                 600                 605
Leu Gly Ser Leu Val Ala Gly Ser Cys Ser Phe Tyr Ser Phe Phe Gly
    610                 615                 620
Glu Pro Thr Val Pro Ala Cys Leu Leu Arg Gln Pro Leu Phe Ser Leu
625                 630                 635                 640
Gly Phe Ala Ile Phe Leu Ser Cys Leu Thr Ile Arg Ser Phe Gln Leu
                645                 650                 655
Val Ile Ile Phe Lys Phe Ser Thr Lys Val Pro Thr Phe Tyr Arg Thr
            660                 665                 670
Trp Ala Gln Asn His Gly Ala Gly Leu Phe Val Ile Val Ser Ser Thr
        675                 680                 685
Val His Leu Leu Ile Cys Leu Thr Trp Leu Val Met Trp Thr Pro Arg
    690                 695                 700
Pro Thr Arg Glu Tyr Gln Arg Phe Pro His Leu Val Ile Leu Glu Cys
705                 710                 715                 720
Thr Glu Val Asn Ser Val Gly Phe Leu Leu Ala Phe Thr His Asn Ile
                725                 730                 735
Leu Leu Ser Ile Ser Thr Phe Val Cys Ser Tyr Leu Gly Lys Glu Leu
            740                 745                 750
Pro Glu Asn Tyr Asn Glu Ala Lys Cys Val Thr Phe Ser Leu Leu Leu
        755                 760                 765
Asn Phe Val Ser Trp Ile Ala Phe Phe Thr Met Ala Ser Ile Tyr Gln
    770                 775                 780
Gly Ser Tyr Leu Pro Ala Val Asn Val Leu Ala Gly Leu Thr Thr Leu
785                 790                 795                 800
Ser Gly Gly Phe Ser Gly Tyr Phe Leu Pro Lys Cys Tyr Val Ile Leu
                805                 810                 815
Cys Arg Pro Glu Leu Asn Asn Thr Glu His Phe Gln Ala Ser Ile Gln
            820                 825                 830
Asp Tyr Thr Arg Arg Cys Gly Thr Thr
        835                 840

&lt;210&gt; 21
&lt;211&gt; 840
&lt;212&gt; PRT
&lt;213&gt; Artificial Sequence

&lt;220&gt;
&lt;223&gt; Description of Artificial Sequence; note =
      synthetic construct

&lt;400&gt; 21
Met Leu Phe Trp Ala Ala His Leu Leu Leu Ser Leu Gln Leu Val Tyr
 1               5                  10                  15
Cys Trp Ala Phe Ser Cys Gln Arg Thr Glu Ser Ser Pro Gly Phe Ser
            20                  25                  30
Leu Pro Gly Asp Phe Leu Leu Ala Gly Leu Phe Ser Leu His Gly Asp
        35                  40                  45
Cys Leu Gln Val Arg His Arg Pro Leu Val Thr Ser Cys Asp Arg Pro
    50                  55                  60
Asp Ser Phe Asn Gly His Gly Tyr His Leu Phe Gln Ala Met Arg Phe
65                  70                  75                  80
Thr Val Glu Glu Ile Asn Asn Ser Ser Ala Leu Leu Pro Asn Ile Thr
```

```
                85                  90                  95
Leu Gly Tyr Glu Leu Tyr Asp Val Cys Ser Glu Ser Ala Asn Val Tyr
            100                 105                 110
Ala Thr Leu Arg Val Leu Ala Leu Gln Gly Pro Arg His Ile Glu Ile
        115                 120                 125
Gln Lys Asp Leu Arg Asn His Ser Ser Lys Val Val Ala Phe Ile Gly
    130                 135                 140
Pro Asp Asn Thr Asp His Ala Val Thr Thr Ala Ala Leu Leu Gly Pro
145                 150                 155                 160
Phe Leu Met Pro Leu Val Ser Tyr Glu Ala Ser Ser Val Val Leu Ser
                165                 170                 175
Ala Lys Arg Lys Phe Pro Ser Phe Leu Arg Thr Val Pro Ser Asp Arg
            180                 185                 190
His Gln Val Glu Val Met Val Gln Leu Leu Gln Ser Phe Gly Trp Val
        195                 200                 205
Trp Ile Ser Leu Ile Gly Ser Tyr Gly Asp Tyr Gly Gln Leu Gly Val
    210                 215                 220
Gln Ala Leu Glu Glu Leu Ala Val Pro Arg Gly Ile Cys Val Ala Phe
225                 230                 235                 240
Lys Asp Ile Val Pro Phe Ser Ala Arg Val Gly Asp Pro Arg Met Gln
                245                 250                 255
Ser Met Met Gln His Leu Ala Gln Ala Arg Thr Thr Val Val Val Val
            260                 265                 270
Phe Ser Asn Arg His Leu Ala Arg Val Phe Phe Arg Ser Val Val Leu
        275                 280                 285
Ala Asn Leu Thr Gly Lys Val Trp Val Ala Ser Glu Asp Trp Ala Ile
    290                 295                 300
Ser Thr Tyr Ile Thr Ser Val Thr Gly Ile Gln Gly Ile Gly Thr Val
305                 310                 315                 320
Leu Gly Val Ala Val Gln Gln Arg Gln Val Pro Gly Leu Lys Glu Phe
                325                 330                 335
Glu Glu Ser Tyr Val Arg Ala Val Thr Ala Ala Pro Ser Ala Cys Pro
            340                 345                 350
Glu Gly Ser Trp Cys Ser Thr Asn Gln Leu Cys Arg Glu Cys His Thr
        355                 360                 365
Phe Thr Thr Arg Asn Met Pro Thr Leu Gly Ala Phe Ser Met Ser Ala
    370                 375                 380
Ala Tyr Arg Val Tyr Glu Ala Val Tyr Ala Val Ala His Gly Leu His
385                 390                 395                 400
Gln Leu Leu Gly Cys Thr Ser Glu Ile Cys Ser Arg Gly Pro Val Tyr
                405                 410                 415
Pro Trp Gln Leu Leu Gln Gln Ile Tyr Lys Val Asn Phe Leu Leu His
            420                 425                 430
Glu Asn Thr Val Ala Phe Asp Asp Asn Gly Asp Thr Leu Gly Tyr Tyr
        435                 440                 445
Asp Ile Ile Ala Trp Asp Trp Asn Gly Pro Glu Trp Thr Phe Glu Ile
    450                 455                 460

Ile Gly Ser Ala Ser Leu Ser Pro Val His Leu Asp Ile Asn Lys Thr
465                 470                 475                 480
Lys Ile Gln Trp His Gly Lys Asn Asn Gln Val Pro Val Ser Val Cys
                485                 490                 495

Thr Thr Asp Cys Leu Ala Gly His His Arg Val Val Val Gly Ser His
            500                 505                 510
His Cys Cys Phe Glu Cys Val Pro Cys Glu Ala Gly Thr Phe Leu Asn
        515                 520                 525
```

```
Met Ser Glu Leu His Ile Cys Gln Pro Cys Gly Thr Glu Glu Trp Ala
    530                 535                 540
Pro Lys Glu Ser Thr Thr Cys Phe Pro Arg Thr Val Glu Phe Leu Glu
545                 550                 555                 560
Leu Arg Glu His Thr Ser Trp Val Leu Leu Ala Ala Asn Thr Leu Leu
                565                 570                 575
Leu Leu Leu Leu Leu Gly Thr Ala Gly Leu Phe Ala Trp His Leu Asp
            580                 585                 590
Thr Pro Val Val Arg Ser Ala Gly Gly Arg Leu Cys Phe Leu Met Leu
        595                 600                 605
Gly Ser Leu Ala Ala Gly Ser Gly Ser Leu Tyr Gly Phe Phe Gly Glu
    610                 615                 620
Pro Thr Arg Pro Ala Cys Leu Leu Arg Gln Ala Leu Phe Ala Leu Gly
625                 630                 635                 640
Phe Thr Ile Phe Leu Ser Cys Leu Thr Val Arg Ser Phe Gln Leu Ile
                645                 650                 655
Ile Ile Phe Lys Phe Ser Thr Lys Val Pro Thr Phe Tyr His Ala Trp
            660                 665                 670
Val Gln Asn His Gly Ala Gly Leu Phe Val Met Ile Ser Ser Ala Ala
        675                 680                 685
Gln Leu Leu Ile Cys Leu Thr Trp Leu Val Val Trp Thr Pro Leu Pro
    690                 695                 700
Ala Arg Glu Tyr Gln Arg Phe Pro His Leu Val Met Leu Glu Cys Thr
705                 710                 715                 720
Glu Thr Asn Ser Leu Gly Phe Ile Leu Ala Phe Leu Tyr Asn Gly Leu
                725                 730                 735
Leu Ser Ile Ser Ala Phe Ala Cys Ser Tyr Leu Gly Lys Asp Leu Pro
            740                 745                 750
Glu Asn Tyr Asn Glu Ala Lys Cys Val Thr Phe Ser Leu Leu Phe Asn
        755                 760                 765
Phe Val Ser Trp Ile Ala Phe Phe Thr Thr Ala Ser Val Tyr Asp Gly
    770                 775                 780
Lys Tyr Leu Pro Ala Ala Asn Met Met Ala Gly Leu Ser Ser Leu Ser
785                 790                 795                 800
Ser Gly Phe Gly Gly Tyr Phe Leu Pro Lys Cys Tyr Val Ile Leu Cys
                805                 810                 815
Arg Pro Asp Leu Asn Ser Thr Glu His Phe Gln Ala Ser Ile Gln Asp
            820                 825                 830
Tyr Thr Arg Arg Cys Gly Ser Thr
        835                 840

<210> 22
<211> 838
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence; note =
      synthetic construct

<400> 22
Met Gly Pro Arg Ala Lys Thr Ile Cys Ser Leu Phe Phe Leu Leu Trp
 1               5                  10                  15
Val Leu Ala Glu Pro Ala Glu Asn Ser Asp Phe Tyr Leu Pro Gly Asp
            20                  25                  30
Tyr Leu Leu Gly Gly Leu Phe Ser Leu His Ala Asn Met Lys Gly Ile
        35                  40                  45
```

```
Val His Leu Asn Phe Leu Gln Val Pro Met Cys Lys Glu Tyr Glu Val
    50                  55                  60
Lys Val Ile Gly Tyr Asn Leu Met Gln Ala Met Arg Phe Ala Val Glu
65                  70                  75                  80
Glu Ile Asn Asn Asp Ser Ser Leu Leu Pro Gly Val Leu Leu Gly Tyr
                85                  90                  95
Glu Ile Val Asp Val Cys Tyr Ile Ser Asn Asn Val Gln Pro Val Leu
            100                 105                 110
Tyr Phe Leu Ala His Glu Asp Asn Leu Leu Pro Ile Gln Glu Asp Tyr
        115                 120                 125
Ser Asn Tyr Ile Ser Arg Val Val Ala Val Ile Gly Pro Asp Asn Ser
    130                 135                 140
Glu Ser Val Met Thr Val Ala Asn Phe Leu Ser Leu Phe Leu Leu Pro
145                 150                 155                 160
Gln Ile Thr Tyr Ser Ala Ile Ser Asp Glu Leu Arg Asp Lys Val Arg
                165                 170                 175
Phe Pro Ala Leu Leu Arg Thr Thr Pro Ser Ala Asp His His Val Glu
            180                 185                 190
Ala Met Val Gln Leu Met Leu His Phe Arg Trp Asn Trp Ile Ile Val
        195                 200                 205
Leu Val Ser Ser Asp Thr Tyr Gly Arg Asp Asn Gly Gln Leu Leu Gly
    210                 215                 220
Glu Arg Val Ala Arg Arg Asp Ile Cys Ile Ala Phe Gln Glu Thr Leu
225                 230                 235                 240
Pro Thr Leu Gln Pro Asn Gln Asn Met Thr Ser Glu Glu Arg Gln Arg
                245                 250                 255
Leu Val Thr Ile Val Asp Lys Leu Gln Gln Ser Thr Ala Arg Val Val
            260                 265                 270
Val Val Phe Ser Pro Asp Leu Thr Leu Tyr His Phe Phe Asn Glu Val
        275                 280                 285
Leu Arg Gln Asn Phe Thr Gly Ala Val Trp Ile Ala Ser Glu Ser Trp
    290                 295                 300
Ala Ile Asp Pro Val Leu His Asn Leu Thr Glu Leu Gly His Leu Gly
305                 310                 315                 320
Thr Phe Leu Gly Ile Thr Ile Gln Ser Val Pro Ile Pro Gly Phe Ser
                325                 330                 335
Glu Phe Arg Glu Trp Gly Pro Gln Ala Gly Pro Pro Pro Leu Ser Arg
            340                 345                 350
Thr Ser Gln Ser Tyr Thr Cys Asn Gln Glu Cys Asp Asn Cys Leu Asn
        355                 360                 365
Ala Thr Leu Ser Phe Asn Thr Ile Leu Arg Leu Ser Gly Glu Arg Val
    370                 375                 380
Val Tyr Ser Val Tyr Ser Ala Val Tyr Ala Val Ala His Ala Leu His
385                 390                 395                 400
Ser Leu Leu Gly Cys Asp Lys Ser Thr Cys Thr Lys Arg Val Val Tyr
                405                 410                 415
Pro Trp Gln Leu Leu Glu Glu Ile Trp Lys Val Asn Phe Thr Leu Leu
            420                 425                 430
Asp His Gln Ile Phe Phe Asp Pro Gln Gly Asp Val Ala Leu His Leu
        435                 440                 445
Glu Ile Val Gln Trp Gln Trp Asp Arg Ser Gln Asn Pro Phe Gln Ser
    450                 455                 460
Val Ala Ser Tyr Tyr Pro Leu Gln Arg Gln Leu Lys Asn Ile Gln Asp
465                 470                 475                 480
Ile Ser Trp His Thr Val Asn Asn Thr Ile Pro Met Ser Met Cys Ser
                485                 490                 495
Lys Arg Cys Gln Ser Gly Gln Lys Lys Lys Pro Val Gly Ile His Val
```

500 505 510
Cys Cys Phe Glu Cys Ile Asp Cys Leu Pro Gly Thr Phe Leu Asn His
515 520 525
Thr Glu Asp Glu Tyr Glu Cys Gln Ala Cys Pro Asn Asn Glu Trp Ser
530 535 540
Tyr Gln Ser Glu Thr Ser Cys Phe Lys Arg Gln Leu Val Phe Leu Glu
545 550 555 560
His Glu Val Pro Thr Ile Val Val Ala Ile Leu Ala Ala Leu Gly Phe
565 570 575
Phe Ser Thr Leu Ala Ile Leu Phe Ile Phe Trp Arg His Phe Gln Thr
580 585 590
Pro Met Val Arg Ser Ala Gly Gly Pro Met Cys Phe Leu Met Leu Val
595 600 605
Pro Leu Leu Leu Ala Phe Gly Met Val Pro Val Tyr Val Gly Pro Pro
610 615 620
Thr Val Phe Ser Cys Phe Cys Arg Gln Ala Phe Phe Thr Val Cys Phe
625 630 635 640
Ser Ile Cys Leu Ser Cys Ile Thr Val Arg Ser Phe Gln Ile Val Cys
645 650 655
Val Phe Lys Met Ala Arg Arg Leu Pro Ser Ala Tyr Ser Phe Trp Met
660 665 670
Arg Tyr His Gly Pro Tyr Val Phe Val Ala Phe Ile Thr Ala Ile Lys
675 680 685
Val Ala Leu Val Val Gly Asn Met Leu Ala Thr Thr Ile Asn Pro Ile
690 695 700
Gly Arg Thr Asp Pro Asp Asp Pro Asn Ile Met Ile Leu Ser Cys His
705 710 715 720
Pro Asn Tyr Arg Asn Gly Leu Leu Phe Asn Thr Ser Met Asp Leu Leu
725 730 735
Leu Ser Val Leu Gly Phe Ser Phe Ala Tyr Met Gly Lys Glu Leu Pro
740 745 750
Thr Asn Tyr Asn Glu Ala Lys Phe Ile Thr Leu Ser Met Thr Phe Ser
755 760 765
Phe Thr Ser Ser Ile Ser Leu Cys Thr Phe Met Ser Val His Asp Gly
770 775 780
Val Leu Val Thr Ile Met Asp Leu Leu Val Thr Val Leu Asn Phe Leu
785 790 795 800
Ala Ile Gly Leu Gly Tyr Phe Gly Pro Lys Cys Tyr Met Ile Leu Phe
805 810 815
Tyr Pro Glu Arg Asn Thr Ser Ala Tyr Phe Asn Ser Met Ile Gln Gly
820 825 830
Tyr Thr Met Arg Lys Ser
835

<210> 23
<211> 843
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence; note =
synthetic construct

<400> 23
Met Gly Pro Gln Ala Arg Thr Leu Cys Leu Leu Ser Leu Leu Leu His
1 5 10 15
Val Leu Pro Lys Pro Gly Lys Leu Val Glu Asn Ser Asp Phe His Leu

```
                20                  25                  30
Ala Gly Asp Tyr Leu Leu Gly Gly Leu Phe Thr Leu His Ala Asn Val
            35                  40                  45
Lys Ser Ile Ser His Leu Ser Tyr Leu Gln Val Pro Lys Cys Asn Glu
        50                  55                  60
Phe Thr Met Lys Val Leu Gly Tyr Asn Leu Met Gln Ala Met Arg Phe
65                  70                  75                  80
Ala Val Glu Glu Ile Asn Asn Cys Ser Ser Leu Leu Pro Gly Val Leu
                85                  90                  95
Leu Gly Tyr Glu Met Val Asp Val Cys Tyr Leu Ser Asn Asn Ile His
            100                 105                 110
Pro Gly Leu Tyr Phe Leu Ala Gln Asp Asp Asp Leu Leu Pro Ile Leu
        115                 120                 125
Lys Asp Tyr Ser Gln Tyr Met Pro His Val Val Ala Val Ile Gly Pro
    130                 135                 140
Asp Asn Ser Glu Ser Ala Ile Thr Val Ser Asn Ile Leu Ser His Phe
145                 150                 155                 160
Leu Ile Pro Gln Ile Thr Tyr Ser Ala Ile Ser Asp Lys Leu Arg Asp
                165                 170                 175
Lys Arg His Phe Pro Ser Met Leu Arg Thr Val Pro Ser Ala Thr His
            180                 185                 190
His Ile Glu Ala Met Val Gln Leu Met Val His Phe Gln Trp Asn Trp
        195                 200                 205
Ile Val Val Leu Val Ser Asp Asp Asp Tyr Gly Arg Glu Asn Ser His
    210                 215                 220
Leu Leu Ser Gln Arg Leu Thr Lys Thr Ser Asp Ile Cys Ile Ala Phe
225                 230                 235                 240
Gln Glu Val Leu Pro Ile Pro Glu Ser Ser Gln Val Met Arg Ser Glu
                245                 250                 255
Glu Gln Arg Gln Leu Asp Asn Ile Leu Asp Lys Leu Arg Arg Thr Ser
            260                 265                 270
Ala Arg Val Val Val Val Phe Ser Pro Glu Leu Ser Leu Tyr Ser Phe
        275                 280                 285
Phe His Glu Val Leu Arg Trp Asn Phe Thr Gly Phe Val Trp Ile Ala
    290                 295                 300
Ser Glu Ser Trp Ala Ile Asp Pro Val Leu His Asn Leu Thr Glu Leu
305                 310                 315                 320
Arg His Thr Gly Thr Phe Leu Gly Val Thr Ile Gln Arg Val Ser Ile
                325                 330                 335
Pro Gly Phe Ser Gln Phe Arg Val Arg Arg Asp Lys Pro Gly Tyr Pro
            340                 345                 350
Val Pro Asn Thr Thr Asn Leu Arg Thr Thr Cys Asn Gln Asp Cys Asp
        355                 360                 365
Ala Cys Leu Asn Thr Thr Lys Ser Phe Asn Asn Ile Leu Ile Leu Ser
    370                 375                 380
Gly Glu Arg Val Val Tyr Ser Val Tyr Ser Ala Val Tyr Ala Val Ala
385                 390                 395                 400
His Ala Leu His Arg Leu Leu Gly Cys Asn Arg Val Arg Cys Thr Lys
                405                 410                 415
Gln Lys Val Tyr Pro Trp Gln Leu Leu Arg Glu Ile Trp His Val Asn
            420                 425                 430
Phe Thr Leu Leu Gly Asn Arg Leu Phe Phe Asp Gln Gln Gly Asp Met
        435                 440                 445
Pro Met Leu Leu Asp Ile Ile Gln Trp Gln Trp Asp Leu Ser Gln Asn
    450                 455                 460
Pro Phe Gln Ser Ile Ala Ser Tyr Ser Pro Thr Ser Lys Arg Leu Thr
465                 470                 475                 480
```

Tyr Ile Asn Asn Val Ser Trp Tyr Thr Pro Asn Asn Thr Val Pro Val
485 490 495
Ser Met Cys Ser Lys Ser Cys Gln Pro Gly Gln Met Lys Lys Ser Val
500 505 510
Gly Leu His Pro Cys Cys Phe Glu Cys Leu Asp Cys Met Pro Gly Thr
515 520 525
Tyr Leu Asn Arg Ser Ala Asp Glu Phe Asn Cys Leu Ser Cys Pro Gly
530 535 540
Ser Met Trp Ser Tyr Lys Asn Asp Ile Thr Cys Phe Gln Arg Arg Pro
545 550 555 560
Thr Phe Leu Glu Trp His Glu Ala Pro Thr Ile Ala Val Ala Leu Leu
565 570 575
Ala Ala Leu Gly Phe Leu Ser Thr Leu Ala Ile Leu Val Ile Phe Trp
580 585 590
Arg His Phe Gln Thr Pro Ile Val Arg Ser Ala Gly Gly Pro Met Cys
595 600 605
Phe Leu Met Leu Thr Leu Leu Leu Val Ala Tyr Met Val Val Pro Val
610 615 620
Tyr Val Gly Pro Pro Lys Val Ser Thr Cys Leu Cys Arg Gln Ala Leu
625 630 635 640
Phe Pro Leu Cys Phe Thr Ile Cys Ile Ser Cys Ile Ala Val Arg Ser
645 650 655
Phe Gln Ile Val Cys Ala Phe Lys Met Ala Ser Arg Phe Pro Arg Ala
660 665 670
Tyr Ser Tyr Trp Val Arg Tyr Gln Gly Pro Tyr Val Ser Met Ala Phe
675 680 685
Ile Thr Val Leu Lys Met Val Ile Val Val Ile Gly Met Leu Ala Thr
690 695 700
Gly Leu Ser Pro Thr Thr Arg Thr Asp Pro Asp Asp Pro Lys Ile Thr
705 710 715 720
Ile Val Ser Cys Asn Pro Asn Tyr Arg Asn Ser Leu Leu Phe Asn Thr
725 730 735
Ser Leu Asp Leu Leu Leu Ser Val Val Gly Phe Ser Phe Ala Tyr Met
740 745 750
Gly Lys Glu Leu Pro Thr Asn Tyr Asn Glu Ala Lys Phe Ile Thr Leu
755 760 765
Ser Met Thr Phe Tyr Phe Thr Ser Ser Val Ser Leu Cys Thr Phe Met
770 775 780
Ser Ala Tyr Ser Gly Val Leu Val Thr Ile Val Asp Leu Leu Val Thr
785 790 795 800
Val Leu Asn Leu Leu Ala Ile Ser Leu Gly Tyr Phe Gly Pro Lys Cys
805 810 815
Tyr Met Ile Leu Phe Tyr Pro Glu Arg Asn Thr Pro Ala Tyr Phe Asn
820 825 830
Ser Met Ile Gln Gly Tyr Thr Met Arg Arg Asp
835 840

<210> 24
<211> 853
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence; note = synthetic construct

<400> 24

Met Leu Gly Pro Ala Val Leu Gly Leu Ser Leu Trp Ala Leu Leu His
1 5 10 15
Pro Gly Thr Gly Ala Pro Leu Cys Leu Ser Gln Gln Leu Arg Met Lys
20 25 30
Gly Asp Tyr Val Leu Gly Gly Leu Phe Pro Leu Gly Glu Ala Glu Glu
35 40 45
Ala Gly Leu Arg Ser Arg Thr Arg Pro Ser Ser Pro Val Cys Thr Arg
50 55 60

Phe Ser Ser Asn Gly Leu Leu Trp Ala Leu Ala Met Lys Met Ala Val
65 70 75 80
Glu Glu Ile Asn Asn Lys Ser Asp Leu Leu Pro Gly Leu Arg Leu Gly
85 90 95
Tyr Asp Leu Phe Asp Thr Cys Ser Glu Pro Val Val Ala Met Lys Pro
100 105 110
Ser Leu Met Phe Leu Ala Lys Ala Gly Ser Arg Asp Ile Ala Ala Tyr
115 120 125
Cys Asn Tyr Thr Gln Tyr Gln Pro Arg Val Leu Ala Val Ile Gly Pro
130 135 140
His Ser Ser Glu Leu Ala Met Val Thr Gly Lys Phe Phe Ser Phe Phe
145 150 155 160
Leu Met Pro Gln Val Ser Tyr Gly Ala Ser Met Glu Leu Leu Ser Ala
165 170 175
Arg Glu Thr Phe Pro Ser Phe Phe Arg Thr Val Pro Ser Asp Arg Val
180 185 190
Gln Leu Thr Ala Ala Ala Glu Leu Leu Gln Glu Phe Gly Trp Asn Trp
195 200 205
Val Ala Ala Leu Gly Ser Asp Asp Glu Tyr Gly Arg Gln Gly Leu Ser
210 215 220
Ile Phe Ser Ala Leu Ala Ala Ala Arg Gly Ile Cys Ile Ala His Glu
225 230 235 240
Gly Leu Val Pro Leu Pro Arg Ala Asp Asp Ser Arg Leu Gly Lys Val
245 250 255
Gln Asp Val Leu His Gln Val Asn Gln Ser Ser Val Gln Val Val Leu
260 265 270
Leu Phe Ala Ser Val His Ala Ala His Ala Leu Phe Asn Tyr Ser Ile
275 280 285
Ser Ser Arg Leu Ser Pro Lys Val Trp Val Ala Ser Glu Ala Trp Leu
290 295 300
Thr Ser Asp Leu Val Met Gly Leu Pro Gly Met Ala Gln Met Gly Thr
305 310 315 320
Val Leu Gly Phe Leu Gln Arg Gly Ala Gln Leu His Glu Phe Pro Gln
325 330 335
Tyr Val Lys Thr His Leu Ala Leu Ala Thr Asp Pro Ala Phe Cys Ser
340 345 350
Ala Leu Gly Glu Arg Glu Gln Gly Leu Glu Glu Asp Val Val Gly Gln
355 360 365
Arg Cys Pro Gln Cys Asp Cys Ile Thr Leu Gln Asn Val Ser Ala Gly
370 375 380
Leu Asn His His Gln Thr Phe Ser Val Tyr Ala Ala Val Tyr Ser Val
385 390 395 400
Ala Gln Ala Leu His Asn Thr Leu Gln Cys Asn Ala Ser Gly Cys Pro
405 410 415
Ala Gln Asp Pro Val Lys Pro Trp Gln Leu Leu Glu Asn Met Tyr Asn
420 425 430

Leu Thr Phe His Val Gly Gly Leu Pro Leu Arg Phe Asp Ser Ser Gly

```
        435                 440                 445
Asn Val Asp Met Glu Tyr Asp Leu Lys Leu Trp Val Trp Gln Gly Ser
    450                 455                 460
Val Pro Arg Leu His Asp Val Gly Arg Phe Asn Gly Ser Leu Arg Thr
465                 470                 475                 480
Glu Arg Leu Lys Ile Arg Trp His Thr Ser Asp Asn Gln Lys Pro Val
                485                 490                 495
Ser Arg Cys Ser Arg Gln Cys Gln Glu Gly Gln Val Arg Arg Val Lys
            500                 505                 510
Gly Phe His Ser Cys Cys Tyr Asp Cys Val Asp Cys Glu Ala Gly Ser
        515                 520                 525
Tyr Arg Gln Asn Pro Asp Asp Ile Ala Cys Thr Phe Cys Gly Gln Asp
    530                 535                 540
Glu Trp Ser Pro Glu Arg Ser Thr Arg Cys Phe Arg Arg Arg Ser Arg
545                 550                 555                 560
Phe Leu Glu Trp Gly Glu Pro Ala Val Leu Ser Leu Leu Leu Leu Leu
                565                 570                 575
Cys Leu Val Leu Gly Leu Thr Leu Ala Ala Leu Gly Leu Phe Val His
            580                 585                 590
Tyr Trp Asp Ser Pro Leu Val Gln Ala Ser Gly Gly Ser Leu Phe Cys
        595                 600                 605
Phe Gly Leu Ile Cys Leu Gly Leu Phe Cys Leu Ser Val Leu Leu Phe
    610                 615                 620
Pro Gly Arg Pro Arg Ser Ala Ser Cys Leu Ala Gln Gln Pro Met Ala
625                 630                 635                 640
His Leu Pro Leu Thr Gly Cys Leu Ser Thr Leu Phe Leu Gln Ala Ala
                645                 650                 655
Glu Ile Phe Val Glu Ser Glu Leu Pro Leu Ser Trp Ala Asn Trp Leu
            660                 665                 670
Cys Ser Tyr Leu Arg Gly Pro Trp Ala Trp Leu Val Val Leu Leu Ala
        675                 680                 685
Thr Leu Val Glu Ala Ala Leu Cys Ala Trp Tyr Leu Met Ala Phe Pro
    690                 695                 700
Pro Glu Val Val Thr Asp Trp Gln Val Leu Pro Thr Glu Val Leu Glu
705                 710                 715                 720
His Cys Arg Met Arg Ser Trp Val Ser Leu Gly Leu Val His Ile Thr
                725                 730                 735
Asn Ala Val Leu Ala Phe Leu Cys Phe Leu Gly Thr Phe Leu Val Gln
            740                 745                 750
Ser Gln Pro Gly Arg Tyr Asn Arg Ala Arg Gly Leu Thr Phe Ala Met
        755                 760                 765
Leu Ala Tyr Phe Ile Ile Trp Val Ser Phe Val Pro Leu Leu Ala Asn
    770                 775                 780
Val Gln Val Ala Tyr Gln Pro Ala Val Gln Met Gly Ala Ile Leu Phe
785                 790                 795                 800
Cys Ala Leu Gly Ile Leu Ala Thr Phe His Leu Pro Lys Cys Tyr Val
                805                 810                 815
Leu Leu Trp Leu Pro Glu Leu Asn Thr Gln Glu Phe Phe Leu Gly Arg
            820                 825                 830
Ser Pro Lys Glu Ala Ser Asp Gly Asn Ser Gly Ser Ser Glu Ala Thr
        835                 840                 845
Arg Gly His Ser Glu
    850
```

<210> 25
<211> 857
<212> PRT

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence; note = synthetic construct

<400> 25
```
Met Pro Gly Leu Ala Ile Leu Gly Leu Ser Leu Ala Ala Phe Leu Glu
 1               5                  10                  15
Leu Gly Met Gly Ser Ser Leu Cys Leu Ser Gln Gln Phe Lys Ala Gln
            20                  25                  30
Gly Asp Tyr Ile Leu Gly Gly Leu Phe Pro Leu Gly Thr Thr Glu Glu
        35                  40                  45
Ala Thr Leu Asn Gln Arg Thr Gln Pro Asn Gly Ile Leu Cys Thr Arg
    50                  55                  60
Phe Ser Pro Leu Gly Leu Phe Leu Ala Met Ala Met Lys Met Ala Val
65                  70                  75                  80
Glu Glu Ile Asn Asn Gly Ser Ala Leu Leu Pro Gly Leu Arg Leu Gly
                85                  90                  95
Tyr Asp Leu Phe Asp Thr Cys Ser Glu Pro Val Val Thr Met Lys Pro
            100                 105                 110
Ser Leu Met Phe Met Ala Lys Val Gly Ser Gln Ser Ile Ala Ala Tyr
        115                 120                 125
Cys Asn Tyr Thr Gln Tyr Gln Pro Arg Val Leu Ala Val Ile Gly Pro
    130                 135                 140
His Ser Ser Glu Leu Ala Leu Ile Thr Gly Lys Phe Phe Ser Phe Phe
145                 150                 155                 160
Leu Met Pro Gln Val Ser Tyr Ser Ala Ser Met Asp Arg Leu Ser Asp
                165                 170                 175
Arg Glu Thr Phe Pro Ser Phe Phe Arg Thr Val Pro Ser Asp Arg Val
            180                 185                 190
Gln Leu Gln Ala Val Val Thr Leu Leu Gln Asn Phe Ser Trp Asn Trp
        195                 200                 205
Val Ala Ala Leu Gly Ser Asp Asp Asp Tyr Gly Arg Glu Gly Leu Ser
    210                 215                 220
Ile Phe Ser Gly Leu Ala Asn Ser Arg Gly Ile Cys Ile Ala His Glu
225                 230                 235                 240
Gly Leu Val Pro Gln His Asp Thr Ser Gly Gln Gln Leu Gly Lys Val
                245                 250                 255
Val Asp Val Leu Arg Gln Val Asn Gln Ser Lys Val Gln Val Val Val
            260                 265                 270
Leu Phe Ala Ser Ala Arg Ala Val Tyr Ser Leu Phe Ser Tyr Ser Ile
        275                 280                 285
Leu His Asp Leu Ser Pro Lys Val Trp Val Ala Ser Glu Ser Trp Leu
    290                 295                 300
Thr Ser Asp Leu Val Met Thr Leu Pro Asn Ile Ala Arg Val Gly Thr
305                 310                 315                 320
Val Leu Gly Phe Leu Gln Arg Gly Ala Leu Leu Pro Glu Phe Ser His
                325                 330                 335
Tyr Val Glu Thr Arg Leu Ala Leu Ala Ala Asp Pro Thr Phe Cys Ala
            340                 345                 350
Ser Leu Lys Ala Glu Leu Asp Leu Glu Glu Arg Val Met Gly Pro Arg
        355                 360                 365
Cys Ser Gln Cys Asp Tyr Ile Met Leu Gln Asn Leu Ser Ser Gly Leu
    370                 375                 380
Met Gln Asn Leu Ser Ala Gly Gln Leu His His Gln Ile Phe Ala Thr
385                 390                 395                 400
```

```
Tyr Ala Ala Val Tyr Ser Val Ala Gln Ala Leu His Asn Thr Leu Gln
                405                 410                 415
Cys Asn Val Ser His Cys His Thr Ser Glu Pro Val Gln Pro Trp Gln
            420                 425                 430
Leu Leu Glu Asn Met Tyr Asn Met Ser Phe Arg Ala Arg Asp Leu Thr
        435                 440                 445
Leu Gln Phe Asp Ala Lys Gly Ser Val Asp Met Glu Tyr Asp Leu Lys
    450                 455                 460
Met Trp Val Trp Gln Ser Pro Thr Pro Val Leu His Thr Val Gly Thr
465                 470                 475                 480
Phe Asn Gly Thr Leu Gln Leu Gln His Ser Lys Met Tyr Trp Pro Gly
                485                 490                 495
Asn Gln Val Pro Val Ser Gln Cys Ser Arg Gln Cys Lys Asp Gly Gln
            500                 505                 510
Val Arg Arg Val Lys Gly Phe His Ser Cys Cys Tyr Asp Cys Val Asp
        515                 520                 525
Cys Lys Ala Gly Ser Tyr Arg Lys His Pro Asp Asp Phe Thr Cys Thr
    530                 535                 540
Pro Cys Gly Lys Asp Gln Trp Ser Pro Glu Lys Ser Thr Thr Cys Leu
545                 550                 555                 560
Pro Arg Arg Pro Lys Phe Leu Glu Trp Gly Glu Pro Ala Val Leu Leu
                565                 570                 575
Leu Leu Leu Leu Leu Ser Leu Ala Leu Gly Leu Val Leu Ala Ala Leu
            580                 585                 590
Gly Leu Phe Val His His Arg Asp Ser Pro Leu Val Gln Ala Ser Gly
        595                 600                 605
Gly Pro Leu Ala Cys Phe Gly Leu Val Cys Leu Gly Leu Val Cys Leu
    610                 615                 620
Ser Val Leu Leu Phe Pro Gly Gln Pro Ser Pro Ala Arg Cys Leu Ala
625                 630                 635                 640
Gln Gln Pro Leu Ser His Leu Pro Leu Thr Gly Cys Leu Ser Thr Leu
                645                 650                 655
Phe Leu Gln Ala Ala Glu Ile Phe Val Glu Ser Glu Leu Pro Leu Ser
            660                 665                 670
Trp Ala Asp Arg Leu Ser Gly Cys Leu Arg Gly Pro Trp Ala Trp Leu
        675                 680                 685
Val Val Leu Leu Ala Met Leu Val Glu Val Ala Leu Cys Thr Trp Tyr
    690                 695                 700
Leu Val Ala Phe Pro Pro Glu Val Val Thr Asp Trp His Met Leu Pro
705                 710                 715                 720
Thr Glu Ala Leu Val His Cys Arg Thr Arg Ser Trp Val Ser Phe Gly
                725                 730                 735
Leu Ala His Ala Thr Asn Ala Thr Leu Ala Phe Leu Cys Phe Leu Gly
            740                 745                 750
Thr Phe Leu Val Arg Ser Gln Pro Gly Arg Tyr Asn Arg Ala Arg Gly
        755                 760                 765
Leu Thr Phe Ala Met Leu Ala Tyr Phe Ile Thr Trp Val Ser Phe Val
    770                 775                 780
Pro Leu Leu Ala Asn Val Gln Val Val Leu Arg Pro Ala Val Gln Met
785                 790                 795                 800
Gly Ala Leu Leu Leu Cys Val Leu Gly Ile Leu Ala Ala Phe His Leu
                805                 810                 815
Pro Arg Cys Tyr Leu Leu Met Arg Gln Pro Gly Leu Asn Thr Pro Glu
            820                 825                 830
Phe Phe Leu Gly Gly Gly Pro Gly Asp Ala Gln Gly Gln Asn Asp Gly
        835                 840                 845
Asn Thr Gly Asn Gln Gly Lys His Glu
```

850 855

<210> 26
<211> 840
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence; note =
synthetic construct

<400> 26
```
Met Gly Pro Arg Ala Lys Thr Ile Cys Ser Leu Phe Phe Leu Leu Trp
 1               5                  10                  15
Val Leu Ala Glu Pro Ala Glu Asn Ser Asp Phe Tyr Leu Pro Gly Asp
            20                  25                  30
Tyr Leu Leu Gly Gly Leu Phe Ser Leu His Ala Asn Met Lys Gly Ile
        35                  40                  45
Val His Leu Asn Phe Leu Gln Val Pro Met Cys Lys Glu Tyr Glu Val
    50                  55                  60
Lys Val Ile Gly Tyr Asn Leu Met Gln Ala Met Arg Phe Ala Val Glu
65                  70                  75                  80
Glu Ile Asn Asn Asp Ser Ser Leu Leu Pro Gly Val Leu Leu Gly Tyr
                85                  90                  95
Glu Ile Val Asp Val Cys Tyr Ile Ser Asn Asn Val Gln Pro Val Leu
            100                 105                 110
Tyr Phe Leu Ala His Glu Asp Asn Leu Leu Pro Ile Gln Glu Asp Tyr
        115                 120                 125
Ser Asn Tyr Ile Ser Arg Val Val Ala Val Ile Gly Pro Asp Asn Ser
    130                 135                 140
Glu Ser Val Met Thr Val Ala Asn Phe Leu Ser Leu Phe Leu Leu Pro
145                 150                 155                 160
Gln Ile Thr Tyr Ser Ala Ile Ser Asp Glu Leu Arg Asp Lys Val Arg
                165                 170                 175
Phe Pro Ala Leu Leu Arg Thr Thr Pro Ser Ala Asp His His Val Glu
            180                 185                 190
Ala Met Val Gln Leu Met Leu His Phe Arg Trp Asn Trp Ile Ile Val
        195                 200                 205
Leu Val Ser Ser Asp Thr Tyr Gly Arg Asp Asn Gly Gln Leu Leu Gly
    210                 215                 220
Glu Arg Val Ala Arg Arg Asp Ile Cys Ile Ala Phe Gln Glu Thr Leu
225                 230                 235                 240
Pro Thr Leu Gln Pro Asn Gln Asn Met Thr Ser Glu Glu Arg Gln Arg
                245                 250                 255
Leu Val Thr Ile Val Asp Lys Leu Gln Gln Ser Thr Ala Arg Val Val
            260                 265                 270
Val Val Phe Ser Pro Asp Leu Thr Leu Tyr His Phe Phe Asn Glu Val
        275                 280                 285
Leu Arg Gln Asn Phe Thr Gly Ala Val Trp Ile Ala Ser Glu Ser Trp
    290                 295                 300
Ala Ile Asp Pro Val Leu His Asn Leu Thr Glu Leu Gly His Leu Gly
305                 310                 315                 320
Thr Phe Leu Gly Ile Thr Ile Gln Ser Val Pro Ile Pro Gly Phe Ser
                325                 330                 335
Glu Phe Arg Glu Trp Gly Pro Gln Ala Gly Pro Pro Pro Leu Ser Arg
            340                 345                 350
Thr Ser Gln Ser Tyr Thr Cys Asn Gln Glu Cys Asp Asn Cys Leu Asn
```

```
        355                 360                 365
Ala Thr Leu Ser Phe Asn Thr Ile Leu Arg Leu Ser Gly Glu Arg Val
    370                 375                 380
Val Tyr Ser Val Tyr Ser Ala Val Tyr Ala Val Ala His Ala Leu His
385                 390                 395                 400
Ser Leu Leu Gly Cys Asp Lys Ser Thr Cys Thr Lys Arg Val Val Tyr
                405                 410                 415
Pro Trp Gln Leu Leu Glu Glu Ile Trp Lys Val Asn Phe Thr Leu Leu
            420                 425                 430
Asp His Gln Ile Phe Phe Asp Pro Gln Gly Asp Val Ala Leu His Leu
        435                 440                 445
Glu Ile Val Gln Trp Gln Trp Asp Arg Ser Gln Asn Pro Phe Gln Ser
    450                 455                 460

Val Ala Ser Tyr Tyr Pro Leu Gln Arg Gln Leu Lys Asn Ile Gln Asp
465                 470                 475                 480
Ile Ser Trp His Thr Val Asn Asn Thr Ile Pro Met Ser Met Cys Ser
                485                 490                 495

Lys Arg Cys Gln Ser Gly Gln Lys Lys Lys Pro Val Gly Ile His Val
            500                 505                 510
Cys Cys Phe Glu Cys Ile Asp Cys Leu Pro Gly Thr Phe Leu Asn His
        515                 520                 525
Thr Glu Asp Glu Tyr Glu Cys Gln Ala Cys Pro Asn Asn Glu Trp Ser
    530                 535                 540
Tyr Gln Ser Glu Thr Ser Cys Phe Lys Arg Gln Leu Val Phe Leu Glu
545                 550                 555                 560
Leu Arg Glu His Thr Ser Trp Val Leu Leu Ala Ala Asn Thr Leu Leu
                565                 570                 575
Leu Leu Leu Leu Leu Gly Thr Ala Gly Leu Phe Ala Trp His Leu Asp
            580                 585                 590
Thr Pro Val Val Arg Ser Ala Gly Gly Arg Leu Cys Phe Leu Met Leu
        595                 600                 605
Gly Ser Leu Ala Ala Gly Ser Gly Ser Leu Tyr Gly Phe Phe Gly Glu
    610                 615                 620
Pro Thr Arg Pro Ala Cys Leu Leu Arg Gln Ala Leu Phe Ala Leu Gly
625                 630                 635                 640
Phe Thr Ile Phe Leu Ser Cys Leu Thr Val Arg Ser Phe Gln Leu Ile
                645                 650                 655
Ile Ile Phe Lys Phe Ser Thr Lys Val Pro Thr Phe Tyr His Ala Trp
            660                 665                 670
Val Gln Asn His Gly Ala Gly Leu Phe Val Met Ile Ser Ser Ala Ala
        675                 680                 685
Gln Leu Leu Ile Cys Leu Thr Trp Leu Val Val Trp Thr Pro Leu Pro
    690                 695                 700
Ala Arg Glu Tyr Gln Arg Phe Pro His Leu Val Met Leu Glu Cys Thr
705                 710                 715                 720
Glu Thr Asn Ser Leu Gly Phe Ile Leu Ala Phe Leu Tyr Asn Gly Leu
                725                 730                 735
Leu Ser Ile Ser Ala Phe Ala Cys Ser Tyr Leu Gly Lys Asp Leu Pro
            740                 745                 750
Glu Asn Tyr Asn Glu Ala Lys Cys Val Thr Phe Ser Leu Leu Phe Asn
        755                 760                 765
Phe Val Ser Trp Ile Ala Phe Phe Thr Thr Ala Ser Val Tyr Asp Gly
    770                 775                 780
Lys Tyr Leu Pro Ala Ala Asn Met Met Ala Gly Leu Ser Ser Leu Ser
785                 790                 795                 800
```

Ser Gly Phe Gly Gly Tyr Phe Leu Pro Lys Cys Tyr Val Ile Leu Cys
805 810 815
Arg Pro Asp Leu Asn Ser Thr Glu His Phe Gln Ala Ser Ile Gln Asp
820 825 830
Tyr Thr Arg Arg Cys Gly Ser Thr
835 840

<210> 27
<211> 840
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence; note =
synthetic construct

<400> 27
Met Leu Leu Cys Thr Ala Arg Leu Val Gly Leu Gln Leu Leu Ile Ser
1 5 10 15
Cys Cys Trp Ala Phe Ala Cys His Ser Thr Glu Ser Ser Pro Asp Phe
20 25 30
Thr Leu Pro Gly Asp Tyr Leu Leu Ala Gly Leu Phe Pro Leu His Ser
35 40 45
Gly Cys Leu Gln Val Arg His Arg Pro Glu Val Thr Leu Cys Asp Arg
50 55 60
Ser Cys Ser Phe Asn Glu His Gly Tyr His Leu Phe Gln Ala Met Arg
65 70 75 80
Leu Gly Val Glu Glu Ile Asn Asn Ser Thr Ala Leu Leu Pro Asn Ile
85 90 95
Thr Leu Gly Tyr Gln Leu Tyr Asp Val Cys Ser Asp Ser Ala Asn Val
100 105 110
Tyr Ala Thr Leu Arg Val Leu Ser Leu Pro Gly Gln His His Ile Glu
115 120 125
Leu Gln Gly Asp Leu Leu His Tyr Ser Pro Thr Val Leu Ala Val Ile
130 135 140
Gly Pro Asp Ser Thr Asn Arg Ala Ala Thr Thr Ala Ala Leu Leu Ser
145 150 155 160
Pro Phe Leu Val Pro Met Ile Ser Tyr Ala Ala Ser Ser Glu Thr Leu
165 170 175
Ser Val Lys Arg Gln Tyr Pro Ser Phe Leu Arg Thr Ile Pro Asn Asp
180 185 190
Lys Tyr Gln Val Glu Thr Met Val Leu Leu Leu Gln Lys Phe Gly Trp
195 200 205
Thr Trp Ile Ser Leu Val Gly Ser Ser Asp Asp Tyr Gly Gln Leu Gly
210 215 220
Val Gln Ala Leu Glu Asn Gln Ala Thr Gly Gln Gly Ile Cys Ile Ala
225 230 235 240
Phe Lys Asp Ile Met Pro Phe Ser Ala Gln Val Gly Asp Glu Arg Met
245 250 255
Gln Cys Leu Met Arg His Leu Ala Gln Ala Gly Ala Thr Val Val Val
260 265 270
Val Phe Ser Ser Arg Gln Leu Ala Arg Val Phe Phe Glu Ser Val Val
275 280 285
Leu Thr Asn Leu Thr Gly Lys Val Trp Val Ala Ser Glu Ala Trp Ala
290 295 300
Leu Ser Arg His Ile Thr Gly Val Pro Gly Ile Gln Arg Ile Gly Met
305 310 315 320

Val Leu Gly Val Ala Ile Gln Lys Arg Ala Val Pro Gly Leu Lys Ala
325 330 335
Phe Glu Glu Ala Tyr Ala Arg Ala Asp Lys Lys Ala Pro Arg Pro Cys
340 345 350
His Lys Gly Ser Trp Cys Ser Ser Asn Gln Leu Cys Arg Glu Cys Gln
355 360 365
Ala Phe Met Ala His Thr Met Pro Lys Leu Lys Ala Phe Ser Met Ser
370 375 380
Ser Ala Tyr Asn Ala Tyr Arg Ala Val Tyr Ala Val Ala His Gly Leu
385 390 395 400
His Gln Leu Leu Gly Cys Ala Ser Gly Ala Cys Ser Arg Gly Arg Val
405 410 415
Tyr Pro Trp Gln Leu Leu Glu Gln Ile His Lys Val His Phe Leu Leu
420 425 430
His Lys Asp Thr Val Ala Phe Asn Asp Asn Arg Asp Pro Leu Ser Ser
435 440 445
Tyr Asn Ile Ile Ala Trp Asp Trp Asn Gly Pro Lys Trp Thr Phe Thr
450 455 460
Val Leu Gly Ser Ser Thr Trp Ser Pro Val Gln Leu Asn Ile Asn Glu
465 470 475 480
Thr Lys Ile Gln Trp His Gly Lys Asp Asn Gln Val Pro Lys Ser Val
485 490 495
Cys Ser Ser Asp Cys Leu Glu Gly His Gln Arg Val Val Thr Gly Phe
500 505 510
His His Cys Cys Phe Glu Cys Val Pro Cys Gly Ala Gly Thr Phe Leu
515 520 525
Asn Lys Ser Asp Leu Tyr Arg Cys Gln Pro Cys Gly Lys Glu Glu Trp
530 535 540
Ala Pro Glu Gly Ser Gln Thr Cys Phe Pro Arg Thr Val Val Phe Leu
545 550 555 560
Glu Trp His Glu Ala Pro Thr Ile Ala Val Ala Leu Leu Ala Ala Leu
565 570 575
Gly Phe Leu Ser Thr Leu Ala Ile Leu Val Ile Phe Trp Arg His Phe
580 585 590
Gln Thr Pro Ile Val Arg Ser Ala Gly Gly Pro Met Cys Phe Leu Met
595 600 605
Leu Thr Leu Leu Leu Val Ala Tyr Met Val Val Pro Val Tyr Val Gly
610 615 620
Pro Pro Lys Val Ser Thr Cys Leu Cys Arg Gln Ala Leu Phe Pro Leu
625 630 635 640
Cys Phe Thr Ile Cys Ile Ser Cys Ile Ala Val Arg Ser Phe Gln Ile
645 650 655
Val Cys Ala Phe Lys Met Ala Ser Arg Phe Pro Arg Ala Tyr Ser Tyr
660 665 670
Trp Val Arg Tyr Gln Gly Pro Tyr Val Ser Met Ala Phe Ile Thr Val
675 680 685
Leu Lys Met Val Ile Val Val Ile Gly Met Leu Ala Thr Gly Leu Ser
690 695 700
Pro Thr Thr Arg Thr Asp Pro Asp Asp Pro Lys Ile Thr Ile Val Ser
705 710 715 720
Cys Asn Pro Asn Tyr Arg Asn Ser Leu Leu Phe Asn Thr Ser Leu Asp
725 730 735
Leu Leu Leu Ser Val Val Gly Phe Ser Phe Ala Tyr Met Gly Lys Glu
740 745 750
Leu Pro Thr Asn Tyr Asn Glu Ala Lys Phe Ile Thr Leu Ser Met Thr
755 760 765
Phe Tyr Phe Thr Ser Ser Val Ser Leu Cys Thr Phe Met Ser Ala Tyr

```
    770                 775                 780
Ser Gly Val Leu Val Thr Ile Val Asp Leu Leu Val Thr Val Leu Asn
785                 790                 795                 800
Leu Leu Ala Ile Ser Leu Gly Tyr Phe Gly Pro Lys Cys Tyr Met Ile
                805                 810                 815
Leu Phe Tyr Pro Glu Arg Asn Thr Pro Ala Tyr Phe Asn Ser Met Ile
            820                 825                 830
Gln Gly Tyr Thr Met Arg Arg Asp
        835                 840

<210> 28
<211> 1123
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence; note =
      synthetic construct

<400> 28
Leu Gln Val Arg His Arg Pro Glu Val Thr Leu Cys Asp Arg Ser Cys
 1               5                  10                  15
Ser Phe Asn Glu His Gly Tyr His Leu Phe Gln Ala Met Arg Leu Gly
            20                  25                  30
Val Glu Glu Ile Asn Asn Ser Thr Ala Leu Leu Pro Asn Ile Thr Leu
        35                  40                  45
Gly Tyr Gln Leu Tyr Asp Val Cys Ser Asp Ser Ala Asn Val Tyr Ala
    50                  55                  60
Thr Leu Arg Val Leu Ser Leu Pro Gly Gln His His Ile Glu Leu Gln
65                  70                  75                  80
Gly Asp Leu Leu His Tyr Ser Pro Thr Val Leu Ala Val Ile Gly Pro
                85                  90                  95
Asp Ser Thr Asn Arg Ala Ala Thr Thr Ala Ala Leu Leu Ser Pro Phe
            100                 105                 110
Leu Val Pro Met Ile Ser Tyr Ala Ala Ser Ser Glu Thr Leu Ser Val
        115                 120                 125
Lys Arg Gln Tyr Pro Ser Phe Leu Arg Thr Ile Pro Asn Asp Lys Tyr
    130                 135                 140
Gln Val Glu Thr Met Val Leu Leu Leu Gln Lys Phe Gly Trp Thr Trp
145                 150                 155                 160
Ile Ser Leu Val Gly Ser Ser Asp Asp Tyr Gly Gln Leu Gly Val Gln
                165                 170                 175
Ala Leu Glu Asn Gln Ala Thr Gly Gln Gly Ile Cys Ile Ala Phe Lys
            180                 185                 190
Asp Ile Met Pro Phe Ser Ala Gln Val Gly Asp Glu Arg Met Gln Cys
        195                 200                 205
Leu Met Arg His Leu Ala Gln Ala Gly Ala Thr Val Val Val Val Phe
    210                 215                 220
Ser Ser Arg Gln Leu Ala Arg Val Phe Phe Glu Ser Val Val Leu Thr
225                 230                 235                 240
Asn Leu Thr Gly Lys Val Trp Val Ala Ser Glu Ala Trp Ala Leu Ser
                245                 250                 255
Arg His Ile Thr Gly Val Pro Gly Ile Gln Arg Ile Gly Met Val Leu
            260                 265                 270
Gly Val Ala Ile Gln Lys Arg Ala Val Pro Gly Leu Lys Ala Phe Glu
        275                 280                 285
Glu Ala Tyr Ala Arg Ala Asp Lys Lys Ala Pro Arg Pro Cys His Lys
```

290 295 300
Gly Ser Trp Cys Ser Ser Asn Gln Leu Cys Arg Glu Cys Gln Ala Phe
305 310 315 320
Met Ala His Thr Met Pro Lys Leu Lys Ala Phe Ser Met Ser Ser Ala
325 330 335
Tyr Asn Ala Tyr Arg Ala Val Tyr Ala Val Ala His Gly Leu His Gln
340 345 350
Leu Leu Gly Cys Ala Ser Gly Ala Cys Ser Arg Gly Arg Val Tyr Pro
355 360 365
Trp Gln Leu Leu Glu Gln Ile His Lys Val His Phe Leu Leu His Lys
370 375 380
Asp Thr Val Ala Phe Asn Asp Asn Arg Asp Pro Leu Ser Ser Tyr Asn
385 390 395 400
Ile Ile Ala Trp Asp Trp Asn Gly Pro Lys Trp Thr Phe Thr Val Leu
405 410 415
Gly Ser Ser Thr Trp Ser Pro Val Gln Leu Asn Ile Asn Glu Thr Lys
420 425 430
Ile Gln Trp His Gly Lys Asp Asn Gln Val Pro Lys Ser Val Cys Ser
435 440 445
Ser Asp Cys Leu Glu Gly His Gln Arg Val Val Thr Gly Phe His His
450 455 460
Cys Cys Phe Glu Cys Val Pro Cys Gly Ala Gly Thr Phe Leu Asn Lys
465 470 475 480
Ser Asp Leu Tyr Arg Cys Gln Pro Cys Gly Lys Glu Glu Trp Ala Pro
485 490 495
Glu Gly Ser Gln Thr Cys Phe Pro Arg Thr Val Val Phe Leu Glu Trp
500 505 510
Ser Asp Ile Glu Ser Ile Ile Ala Ile Ala Phe Ser Cys Leu Gly Ile
515 520 525
Leu Val Thr Leu Phe Val Thr Leu Ile Phe Val Leu Tyr Arg Asp Thr
530 535 540
Pro Val Val Lys Ser Ser Ser Arg Glu Leu Cys Tyr Ile Ile Leu Ala
545 550 555 560
Gly Ile Phe Leu Gly Tyr Val Cys Pro Phe Thr Leu Ile Ala Lys Pro
565 570 575
Thr Thr Thr Ser Cys Tyr Leu Gln Arg Leu Leu Val Gly Leu Ser Ser
580 585 590
Ala Met Cys Tyr Ser Ala Leu Val Thr Lys Thr Asn Arg Ile Ala Arg
595 600 605
Ile Leu Ala Gly Ser Lys Lys Lys Ile Cys Thr Arg Lys Pro Arg Phe
610 615 620
Met Ser Ala Trp Ala Gln Val Ile Ile Ala Ser Ile Leu Ile Ser Val
625 630 635 640
Gln Leu Thr Leu Val Val Thr Leu Ile Ile Met Glu Pro Pro Met Pro
645 650 655
Ile Leu Ser Tyr Pro Ser Ile Lys Glu Val Tyr Leu Ile Cys Asn Thr
660 665 670
Ser Asn Leu Gly Val Val Ala Pro Val Gly Tyr Asn Gly Leu Leu Ile
675 680 685
Met Ser Cys Thr Tyr Tyr Ala Phe Lys Thr Arg Asn Val Pro Ala Asn
690 695 700
Phe Asn Glu Ala Lys Tyr Ile Ala Phe Thr Met Tyr Thr Thr Cys Ile
705 710 715 720
Ile Trp Leu Ala Phe Val Pro Ile Tyr Phe Gly Ser Asn Tyr Lys Ile
725 730 735
Ile Thr Thr Cys Phe Ala Val Ser Leu Ser Val Thr Val Ala Leu Gly
740 745 750

```
Cys Met Phe Thr Pro Lys Met Tyr Ile Ile Ile Ala Lys Pro Glu Arg
        755                 760                 765
Asn Val Arg Ser Ala Phe Thr Thr Ser Asp Val Val Arg Met His Val
    770                 775                 780
Gly Asp Gly Lys Leu Pro Cys Arg Ser Asn Thr Phe Leu Asn Ile Phe
785                 790                 795                 800
Arg Arg Lys Lys Pro Gly Ala Gly Asn Ala Asn Ser Asn Gly Lys Ser
                805                 810                 815
Val Ser Trp Ser Glu Pro Gly Gly Arg Gln Ala Pro Lys Gly Gln His
            820                 825                 830
Val Trp Gln Arg Leu Ser Val His Val Lys Thr Asn Glu Thr Ala Cys
        835                 840                 845
Asn Gln Thr Ala Val Ile Lys Pro Leu Thr Lys Ser Tyr Gln Gly Ser
    850                 855                 860
Gly Lys Ser Leu Thr Phe Ser Asp Ala Ser Thr Lys Thr Leu Tyr Asn
865                 870                 875                 880
Val Glu Glu Glu Asp Asn Thr Pro Ser Ala His Phe Ser Pro Pro Ser
                885                 890                 895
Ser Pro Ser Met Val Val His Arg Arg Gly Pro Pro Val Ala Thr Thr
            900                 905                 910
Pro Pro Leu Pro Pro His Leu Thr Ala Glu Glu Thr Pro Leu Phe Leu
        915                 920                 925
Ala Asp Ser Val Ile Pro Lys Gly Leu Pro Pro Pro Leu Pro Gln Gln
    930                 935                 940
Gln Pro Gln Gln Pro Pro Pro Gln Gln Pro Pro Gln Gln Pro Lys Ser
945                 950                 955                 960
Leu Met Asp Gln Leu Gln Gly Val Val Thr Asn Phe Gly Ser Gly Ile
                965                 970                 975
Pro Asp Phe His Ala Val Leu Ala Gly Pro Gly Thr Pro Gly Asn Ser
            980                 985                 990
Leu Arg Ser Leu Tyr Pro Pro Pro Pro Pro Pro Gln His Leu Gln Met
        995                 1000                1005
Leu Pro Leu His Leu Ser Thr Phe Gln Glu Glu Ser Ile Ser Pro Pro
    1010                1015                1020
Gly Glu Asp Ile Asp Asp Asp Ser Glu Arg Phe Lys Leu Leu Gln Glu
1025                1030                1035                1040
Phe Val Tyr Glu Arg Glu Gly Asn Thr Glu Glu Asp Glu Leu Glu Glu
                1045                1050                1055
Glu Glu Asp Leu Pro Thr Ala Ser Lys Leu Thr Pro Glu Asp Ser Pro
            1060                1065                1070
Ala Leu Thr Pro Pro Ser Pro Phe Arg Asp Ser Val Ala Ser Gly Ser
        1075                1080                1085
Ser Val Pro Ser Ser Pro Val Ser Glu Ser Val Leu Cys Thr Pro Pro
    1090                1095                1100
Asn Val Thr Tyr Ala Ser Val Ile Leu Arg Asp Tyr Lys Gln Ser Ser
1105                1110                1115                1120
Ser Thr Leu
```

<210> 29
<211> 1172
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence; note =
synthetic construct

<400> 29

```
Met Gly Pro Arg Ala Lys Thr Ile Cys Ser Leu Phe Phe Leu Leu Trp
1               5                   10                  15
Val Leu Ala Glu Pro Ala Glu Asn Ser Asp Phe Tyr Leu Pro Gly Asp
            20                  25                  30
Tyr Leu Leu Gly Gly Leu Phe Ser Leu His Ala Asn Met Lys Gly Ile
        35                  40                  45
Val His Leu Asn Phe Leu Gln Val Pro Met Cys Lys Glu Tyr Glu Val
    50                  55                  60
Lys Val Ile Gly Tyr Asn Leu Met Gln Ala Met Arg Phe Ala Val Glu
65                  70                  75                  80
Glu Ile Asn Asn Asp Ser Ser Leu Leu Pro Gly Val Leu Leu Gly Tyr
                85                  90                  95
Glu Ile Val Asp Val Cys Tyr Ile Ser Asn Asn Val Gln Pro Val Leu
            100                 105                 110
Tyr Phe Leu Ala His Glu Asp Asn Leu Leu Pro Ile Gln Glu Asp Tyr
        115                 120                 125
Ser Asn Tyr Ile Ser Arg Val Val Ala Val Ile Gly Pro Asp Asn Ser
    130                 135                 140
Glu Ser Val Met Thr Val Ala Asn Phe Leu Ser Leu Phe Leu Leu Pro
145                 150                 155                 160
Gln Ile Thr Tyr Ser Ala Ile Ser Asp Glu Leu Arg Asp Lys Val Arg
                165                 170                 175
Phe Pro Ala Leu Leu Arg Thr Thr Pro Ser Ala Asp His His Val Glu
            180                 185                 190
Ala Met Val Gln Leu Met Leu His Phe Arg Trp Asn Trp Ile Ile Val
        195                 200                 205
Leu Val Ser Ser Asp Thr Tyr Gly Arg Asp Asn Gly Gln Leu Leu Gly
    210                 215                 220
Glu Arg Val Ala Arg Arg Asp Ile Cys Ile Ala Phe Gln Glu Thr Leu
225                 230                 235                 240
Pro Thr Leu Gln Pro Asn Gln Asn Met Thr Ser Glu Glu Arg Gln Arg
                245                 250                 255
Leu Val Thr Ile Val Asp Lys Leu Gln Gln Ser Thr Ala Arg Val Val
            260                 265                 270
Val Val Phe Ser Pro Asp Leu Thr Leu Tyr His Phe Phe Asn Glu Val
        275                 280                 285
Leu Arg Gln Asn Phe Thr Gly Ala Val Trp Ile Ala Ser Glu Ser Trp
    290                 295                 300
Ala Ile Asp Pro Val Leu His Asn Leu Thr Glu Leu Gly His Leu Gly
305                 310                 315                 320
Thr Phe Leu Gly Ile Thr Ile Gln Ser Val Pro Ile Pro Gly Phe Ser
                325                 330                 335
Glu Phe Arg Glu Trp Gly Pro Gln Ala Gly Pro Pro Pro Leu Ser Arg
            340                 345                 350
Thr Ser Gln Ser Tyr Thr Cys Asn Gln Glu Cys Asp Asn Cys Leu Asn
        355                 360                 365
Ala Thr Leu Ser Phe Asn Thr Ile Leu Arg Leu Ser Gly Glu Arg Val
    370                 375                 380
Val Tyr Ser Val Tyr Ser Ala Val Tyr Ala Val Ala His Ala Leu His
385                 390                 395                 400
Ser Leu Leu Gly Cys Asp Lys Ser Thr Cys Thr Lys Arg Val Val Tyr
                405                 410                 415
Pro Trp Gln Leu Leu Glu Glu Ile Trp Lys Val Asn Phe Thr Leu Leu
            420                 425                 430
Asp His Gln Ile Phe Phe Asp Pro Gln Gly Asp Val Ala Leu His Leu
```

```
        435                 440                 445
Glu Ile Val Gln Trp Gln Trp Asp Arg Ser Gln Asn Pro Phe Gln Ser
    450                 455                 460
Val Ala Ser Tyr Tyr Pro Leu Gln Arg Gln Leu Lys Asn Ile Gln Asp
465                 470                 475                 480
Ile Ser Trp His Thr Val Asn Asn Thr Ile Pro Met Ser Met Cys Ser
                485                 490                 495
Lys Arg Cys Gln Ser Gly Gln Lys Lys Lys Pro Val Gly Ile His Val
            500                 505                 510
Cys Cys Phe Glu Cys Ile Asp Cys Leu Pro Gly Thr Phe Leu Asn His
        515                 520                 525
Thr Glu Asp Glu Tyr Glu Cys Gln Ala Cys Pro Asn Asn Glu Trp Ser
    530                 535                 540
Tyr Gln Ser Glu Thr Ser Cys Phe Lys Arg Gln Leu Val Phe Leu Glu
545                 550                 555                 560
Trp Ser Asp Ile Glu Ser Ile Ile Ala Ile Ala Phe Ser Cys Leu Gly
                565                 570                 575
Ile Leu Val Thr Leu Phe Val Thr Leu Ile Phe Val Leu Tyr Arg Asp
            580                 585                 590
Thr Pro Val Val Lys Ser Ser Ser Arg Glu Leu Cys Tyr Ile Ile Leu
        595                 600                 605
Ala Gly Ile Phe Leu Gly Tyr Val Cys Pro Phe Thr Leu Ile Ala Lys
    610                 615                 620
Pro Thr Thr Thr Ser Cys Tyr Leu Gln Arg Leu Leu Val Gly Leu Ser
625                 630                 635                 640
Ser Ala Met Cys Tyr Ser Ala Leu Val Thr Lys Thr Asn Arg Ile Ala
                645                 650                 655
Arg Ile Leu Ala Gly Ser Lys Lys Lys Ile Cys Thr Arg Lys Pro Arg
            660                 665                 670
Phe Met Ser Ala Trp Ala Gln Val Ile Ile Ala Ser Ile Leu Ile Ser
        675                 680                 685
Val Gln Leu Thr Leu Val Val Thr Leu Ile Ile Met Glu Pro Pro Met
    690                 695                 700
Pro Ile Leu Ser Tyr Pro Ser Ile Lys Glu Val Tyr Leu Ile Cys Asn
705                 710                 715                 720
Thr Ser Asn Leu Gly Val Val Ala Pro Val Gly Tyr Asn Gly Leu Leu
                725                 730                 735
Ile Met Ser Cys Thr Tyr Tyr Ala Phe Lys Thr Arg Asn Val Pro Ala
            740                 745                 750
Asn Phe Asn Glu Ala Lys Tyr Ile Ala Phe Thr Met Tyr Thr Thr Cys
        755                 760                 765
Ile Ile Trp Leu Ala Phe Val Pro Ile Tyr Phe Gly Ser Asn Tyr Lys
    770                 775                 780
Ile Ile Thr Thr Cys Phe Ala Val Ser Leu Ser Val Thr Val Ala Leu
785                 790                 795                 800
Gly Cys Met Phe Thr Pro Lys Met Tyr Ile Ile Ile Ala Lys Pro Glu
                805                 810                 815
Arg Asn Val Arg Ser Ala Phe Thr Thr Ser Asp Val Val Arg Met His
            820                 825                 830
Val Gly Asp Gly Lys Leu Pro Cys Arg Ser Asn Thr Phe Leu Asn Ile
        835                 840                 845
Phe Arg Arg Lys Lys Pro Gly Ala Gly Asn Ala Asn Ser Asn Gly Lys
    850                 855                 860
Ser Val Ser Trp Ser Glu Pro Gly Gly Arg Gln Ala Pro Lys Gly Gln
865                 870                 875                 880
His Val Trp Gln Arg Leu Ser Val His Val Lys Thr Asn Glu Thr Ala
                885                 890                 895
```

Cys Asn Gln Thr Ala Val Ile Lys Pro Leu Thr Lys Ser Tyr Gln Gly
900 905 910
Ser Gly Lys Ser Leu Thr Phe Ser Asp Ala Ser Thr Lys Thr Leu Tyr
915 920 925
Asn Val Glu Glu Glu Asp Asn Thr Pro Ser Ala His Phe Ser Pro Pro
930 935 940
Ser Ser Pro Ser Met Val Val His Arg Arg Gly Pro Pro Val Ala Thr
945 950 955 960
Thr Pro Pro Leu Pro Pro His Leu Thr Ala Glu Glu Thr Pro Leu Phe
965 970 975
Leu Ala Asp Ser Val Ile Pro Lys Gly Leu Pro Pro Pro Leu Pro Gln
980 985 990
Gln Gln Pro Gln Gln Pro Pro Pro Gln Gln Pro Pro Gln Gln Pro Lys
995 1000 1005
Ser Leu Met Asp Gln Leu Gln Gly Val Val Thr Asn Phe Gly Ser Gly
1010 1015 1020
Ile Pro Asp Phe His Ala Val Leu Ala Gly Pro Gly Thr Pro Gly Asn
1025 1030 1035 1040
Ser Leu Arg Ser Leu Tyr Pro Pro Pro Pro Pro Pro Gln His Leu Gln
1045 1050 1055
Met Leu Pro Leu His Leu Ser Thr Phe Gln Glu Glu Ser Ile Ser Pro
1060 1065 1070
Pro Gly Glu Asp Ile Asp Asp Asp Ser Glu Arg Phe Lys Leu Leu Gln
1075 1080 1085
Glu Phe Val Tyr Glu Arg Glu Gly Asn Thr Glu Glu Asp Glu Leu Glu
1090 1095 1100
Glu Glu Glu Asp Leu Pro Thr Ala Ser Lys Leu Thr Pro Glu Asp Ser
1105 1110 1115 1120
Pro Ala Leu Thr Pro Pro Ser Pro Phe Arg Asp Ser Val Ala Ser Gly
1125 1130 1135
Ser Ser Val Pro Ser Ser Pro Val Ser Glu Ser Val Leu Cys Thr Pro
1140 1145 1150
Pro Asn Val Thr Tyr Ala Ser Val Ile Leu Arg Asp Tyr Lys Gln Ser
1155 1160 1165
Ser Ser Thr Leu
1170

<210> 30
<211> 1175
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence; note =
synthetic construct

<400> 30
Met Leu Gly Pro Ala Val Leu Gly Leu Ser Leu Trp Ala Leu Leu His
1 5 10 15
Pro Gly Thr Gly Ala Pro Leu Cys Leu Ser Gln Gln Leu Arg Met Lys
20 25 30
Gly Asp Tyr Val Leu Gly Gly Leu Phe Pro Leu Gly Glu Ala Glu Glu
35 40 45
Ala Gly Leu Arg Ser Arg Thr Arg Pro Ser Ser Pro Val Cys Thr Arg
50 55 60
Phe Ser Ser Asn Gly Leu Leu Trp Ala Leu Ala Met Lys Met Ala Val
65 70 75 80

```
Glu Glu Ile Asn Asn Lys Ser Asp Leu Leu Pro Gly Leu Arg Leu Gly
                85                  90                  95

Tyr Asp Leu Phe Asp Thr Cys Ser Glu Pro Val Val Ala Met Lys Pro
            100                 105                 110
Ser Leu Met Phe Leu Ala Lys Ala Gly Ser Arg Asp Ile Ala Ala Tyr
        115                 120                 125
Cys Asn Tyr Thr Gln Tyr Gln Pro Arg Val Leu Ala Val Ile Gly Pro
    130                 135                 140
His Ser Ser Glu Leu Ala Met Val Thr Gly Lys Phe Phe Ser Phe Phe
145                 150                 155                 160
Leu Met Pro Gln Val Ser Tyr Gly Ala Ser Met Glu Leu Leu Ser Ala
                165                 170                 175
Arg Glu Thr Phe Pro Ser Phe Phe Arg Thr Val Pro Ser Asp Arg Val
            180                 185                 190
Gln Leu Thr Ala Ala Ala Glu Leu Leu Gln Glu Phe Gly Trp Asn Trp
        195                 200                 205
Val Ala Ala Leu Gly Ser Asp Asp Glu Tyr Gly Arg Gln Gly Leu Ser
    210                 215                 220
Ile Phe Ser Ala Leu Ala Ala Ala Arg Gly Ile Cys Ile Ala His Glu
225                 230                 235                 240
Gly Leu Val Pro Leu Pro Arg Ala Asp Asp Ser Arg Leu Gly Lys Val
                245                 250                 255
Gln Asp Val Leu His Gln Val Asn Gln Ser Ser Val Gln Val Val Leu
            260                 265                 270
Leu Phe Ala Ser Val His Ala Ala His Ala Leu Phe Asn Tyr Ser Ile
        275                 280                 285
Ser Ser Arg Leu Ser Pro Lys Val Trp Val Ala Ser Glu Ala Trp Leu
    290                 295                 300
Thr Ser Asp Leu Val Met Gly Leu Pro Gly Met Ala Gln Met Gly Thr
305                 310                 315                 320
Val Leu Gly Phe Leu Gln Arg Gly Ala Gln Leu His Glu Phe Pro Gln
                325                 330                 335
Tyr Val Lys Thr His Leu Ala Leu Ala Thr Asp Pro Ala Phe Cys Ser
            340                 345                 350
Ala Leu Gly Glu Arg Glu Gln Gly Leu Glu Glu Asp Val Val Gly Gln
        355                 360                 365
Arg Cys Pro Gln Cys Asp Cys Ile Thr Leu Gln Asn Val Ser Ala Gly
    370                 375                 380
Leu Asn His His Gln Thr Phe Ser Val Tyr Ala Ala Val Tyr Ser Val
385                 390                 395                 400
Ala Gln Ala Leu His Asn Thr Leu Gln Cys Asn Ala Ser Gly Cys Pro
                405                 410                 415
Ala Gln Asp Pro Val Lys Pro Trp Gln Leu Leu Glu Asn Met Tyr Asn
            420                 425                 430
Leu Thr Phe His Val Gly Gly Leu Pro Leu Arg Phe Asp Ser Ser Gly
        435                 440                 445
Asn Val Asp Met Glu Tyr Asp Leu Lys Leu Trp Val Trp Gln Gly Ser
    450                 455                 460
Val Pro Arg Leu His Asp Val Gly Arg Phe Asn Gly Ser Leu Arg Thr
465                 470                 475                 480
Glu Arg Leu Lys Ile Arg Trp His Thr Ser Asp Asn Gln Lys Pro Val
                485                 490                 495
Ser Arg Cys Ser Arg Gln Cys Gln Glu Gly Gln Val Arg Arg Val Lys
            500                 505                 510
Gly Phe His Ser Cys Cys Tyr Asp Cys Val Asp Cys Glu Ala Gly Ser
        515                 520                 525
```

Tyr Arg Gln Asn Pro Asp Asp Ile Ala Cys Thr Phe Cys Gly Gln Asp
530 535 540
Glu Trp Ser Pro Glu Arg Ser Thr Arg Cys Phe Arg Arg Arg Ser Arg
545 550 555 560
Phe Leu Glu Trp Ser Asp Ile Glu Ser Ile Ile Ala Ile Ala Phe Ser
565 570 575
Cys Leu Gly Ile Leu Val Thr Leu Phe Val Thr Leu Ile Phe Val Leu
580 585 590
Tyr Arg Asp Thr Pro Val Val Lys Ser Ser Ser Arg Glu Leu Cys Tyr
595 600 605
Ile Ile Leu Ala Gly Ile Phe Leu Gly Tyr Val Cys Pro Phe Thr Leu
610 615 620
Ile Ala Lys Pro Thr Thr Thr Ser Cys Tyr Leu Gln Arg Leu Leu Val
625 630 635 640
Gly Leu Ser Ser Ala Met Cys Tyr Ser Ala Leu Val Thr Lys Thr Asn
645 650 655
Arg Ile Ala Arg Ile Leu Ala Gly Ser Lys Lys Lys Ile Cys Thr Arg
660 665 670
Lys Pro Arg Phe Met Ser Ala Trp Ala Gln Val Ile Ile Ala Ser Ile
675 680 685
Leu Ile Ser Val Gln Leu Thr Leu Val Val Thr Leu Ile Ile Met Glu
690 695 700
Pro Pro Met Pro Ile Leu Ser Tyr Pro Ser Ile Lys Glu Val Tyr Leu
705 710 715 720
Ile Cys Asn Thr Ser Asn Leu Gly Val Val Ala Pro Val Gly Tyr Asn
725 730 735
Gly Leu Leu Ile Met Ser Cys Thr Tyr Tyr Ala Phe Lys Thr Arg Asn
740 745 750
Val Pro Ala Asn Phe Asn Glu Ala Lys Tyr Ile Ala Phe Thr Met Tyr
755 760 765

Thr Thr Cys Ile Ile Trp Leu Ala Phe Val Pro Ile Tyr Phe Gly Ser
770 775 780
Asn Tyr Lys Ile Ile Thr Thr Cys Phe Ala Val Ser Leu Ser Val Thr
785 790 795 800
Val Ala Leu Gly Cys Met Phe Thr Pro Lys Met Tyr Ile Ile Ile Ala
805 810 815
Lys Pro Glu Arg Asn Val Arg Ser Ala Phe Thr Thr Ser Asp Val Val
820 825 830
Arg Met His Val Gly Asp Gly Lys Leu Pro Cys Arg Ser Asn Thr Phe
835 840 845
Leu Asn Ile Phe Arg Arg Lys Lys Pro Gly Ala Gly Asn Ala Asn Ser
850 855 860
Asn Gly Lys Ser Val Ser Trp Ser Glu Pro Gly Gly Arg Gln Ala Pro
865 870 875 880
Lys Gly Gln His Val Trp Gln Arg Leu Ser Val His Val Lys Thr Asn
885 890 895
Glu Thr Ala Cys Asn Gln Thr Ala Val Ile Lys Pro Leu Thr Lys Ser
900 905 910
Tyr Gln Gly Ser Gly Lys Ser Leu Thr Phe Ser Asp Ala Ser Thr Lys
915 920 925
Thr Leu Tyr Asn Val Glu Glu Glu Asp Asn Thr Pro Ser Ala His Phe
930 935 940
Ser Pro Pro Ser Ser Pro Ser Met Val Val His Arg Arg Gly Pro Pro
945 950 955 960
Val Ala Thr Thr Pro Pro Leu Pro Pro His Leu Thr Ala Glu Glu Thr
965 970 975

Pro Leu Phe Leu Ala Asp Ser Val Ile Pro Lys Gly Leu Pro Pro Pro
980 985 990
Leu Pro Gln Gln Gln Pro Gln Gln Pro Pro Pro Gln Gln Pro Pro Gln
995 1000 1005
Gln Pro Lys Ser Leu Met Asp Gln Leu Gln Gly Val Val Thr Asn Phe
1010 1015 1020
Gly Ser Gly Ile Pro Asp Phe His Ala Val Leu Ala Gly Pro Gly Thr
1025 1030 1035 1040
Pro Gly Asn Ser Leu Arg Ser Leu Tyr Pro Pro Pro Pro Pro Pro Gln
1045 1050 1055
His Leu Gln Met Leu Pro Leu His Leu Ser Thr Phe Gln Glu Glu Ser
1060 1065 1070
Ile Ser Pro Pro Gly Glu Asp Ile Asp Asp Asp Ser Glu Arg Phe Lys
1075 1080 1085
Leu Leu Gln Glu Phe Val Tyr Glu Arg Glu Gly Asn Thr Glu Glu Asp
1090 1095 1100
Glu Leu Glu Glu Glu Glu Asp Leu Pro Thr Ala Ser Lys Leu Thr Pro
1105 1110 1115 1120
Glu Asp Ser Pro Ala Leu Thr Pro Pro Ser Pro Phe Arg Asp Ser Val
1125 1130 1135
Ala Ser Gly Ser Ser Val Pro Ser Ser Pro Val Ser Glu Ser Val Leu
1140 1145 1150
Cys Thr Pro Pro Asn Val Thr Tyr Ala Ser Val Ile Leu Arg Asp Tyr
1155 1160 1165
Lys Gln Ser Ser Ser Thr Leu
1170 1175

<210> 31
<211> 867
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence; note =
synthetic construct

<400> 31
Met Val Arg Leu Leu Leu Ile Phe Phe Pro Met Ile Phe Leu Glu Met
1 5 10 15
Ser Ile Leu Pro Arg Met Pro Asp Arg Lys Val Leu Leu Ala Gly Ala
20 25 30
Ser Ser Gln Arg Ser Val Ala Arg Met Asp Gly Asp Val Ile Ile Gly
35 40 45
Ala Leu Phe Ser Val His His Gln Pro Pro Ala Glu Lys Val Pro Glu
50 55 60
Arg Lys Cys Gly Glu Ile Arg Glu Gln Tyr Gly Ile Gln Arg Val Glu
65 70 75 80
Ala Met Phe His Thr Leu Asp Lys Ile Asn Ala Asp Pro Val Leu Leu
85 90 95
Pro Asn Ile Thr Leu Gly Ser Glu Ile Arg Asp Ser Cys Trp His Ser
100 105 110
Ser Val Ala Leu Glu Gln Ser Ile Glu Phe Ile Arg Asp Ser Leu Ile
115 120 125
Ser Ile Arg Asp Glu Lys Asp Gly Leu Asn Arg Cys Leu Pro Asp Gly
130 135 140
Gln Thr Leu Pro Pro Gly Arg Thr Lys Lys Pro Ile Ala Gly Val Ile
145 150 155 160

Gly Pro Gly Ser Ser Ser Val Ala Ile Gln Val Gln Asn Leu Leu Gln
165 170 175
Leu Phe Asp Ile Pro Gln Ile Ala Tyr Ser Ala Thr Ser Ile Asp Leu
180 185 190
Ser Asp Lys Thr Leu Tyr Lys Tyr Phe Leu Arg Val Val Pro Ser Asp
195 200 205
Thr Leu Gln Ala Arg Ala Met Leu Asp Ile Val Lys Arg Tyr Asn Trp
210 215 220
Thr Tyr Val Ser Ala Val His Thr Glu Gly Asn Tyr Gly Glu Ser Gly
225 230 235 240
Met Asp Ala Phe Lys Glu Leu Ala Ala Gln Glu Gly Leu Cys Ile Ala
245 250 255
His Ser Asp Lys Ile Tyr Ser Asn Ala Gly Glu Lys Ser Phe Asp Arg
260 265 270
Leu Leu Arg Lys Leu Arg Glu Arg Leu Pro Lys Ala Arg Val Val Val
275 280 285
Cys Phe Cys Glu Gly Met Thr Val Arg Gly Leu Leu Ser Ala Met Arg
290 295 300
Arg Leu Gly Val Val Gly Glu Phe Ser Leu Ile Gly Ser Asp Gly Trp
305 310 315 320
Ala Asp Arg Asp Glu Val Ile Glu Gly Tyr Glu Val Glu Ala Asn Gly
325 330 335
Gly Ile Thr Ile Lys Leu Gln Ser Pro Glu Val Arg Ser Phe Asp Asp
340 345 350
Tyr Phe Leu Lys Leu Arg Leu Asp Thr Asn Thr Arg Asn Pro Trp Phe
355 360 365
Pro Glu Phe Trp Gln His Arg Phe Gln Cys Arg Leu Pro Gly His Leu
370 375 380
Leu Glu Asn Pro Asn Phe Lys Lys Val Cys Thr Gly Asn Glu Ser Leu
385 390 395 400
Glu Glu Asn Tyr Val Gln Asp Ser Lys Met Gly Phe Val Ile Asn Ala
405 410 415
Ile Tyr Ala Met Ala His Gly Leu Gln Asn Met His His Ala Leu Cys
420 425 430
Pro Gly His Val Gly Leu Cys Asp Ala Met Lys Pro Ile Asp Gly Arg
435 440 445
Lys Leu Leu Asp Phe Leu Ile Lys Ser Ser Phe Val Gly Val Ser Gly
450 455 460
Glu Glu Val Trp Phe Asp Glu Lys Gly Asp Ala Pro Gly Arg Tyr Asp
465 470 475 480
Ile Met Asn Leu Gln Tyr Thr Glu Ala Asn Arg Tyr Asp Tyr Val His
485 490 495
Val Gly Thr Trp His Glu Gly Val Leu Asn Ile Asp Asp Tyr Lys Ile
500 505 510
Gln Met Asn Lys Ser Gly Met Val Arg Ser Val Cys Ser Glu Pro Cys
515 520 525
Leu Lys Gly Gln Ile Lys Val Ile Arg Lys Gly Glu Val Ser Cys Cys
530 535 540
Trp Ile Cys Thr Ala Cys Lys Glu Asn Glu Phe Val Gln Asp Glu Phe
545 550 555 560
Thr Cys Arg Ala Cys Asp Leu Gly Trp Trp Pro Asn Ala Glu Leu Thr
565 570 575
Gly Cys Glu Pro Ile Pro Val Arg Tyr Leu Glu Leu Arg Glu His Thr
580 585 590
Ser Trp Val Leu Leu Ala Ala Asn Thr Leu Leu Leu Leu Leu Leu
595 600 605
Gly Thr Ala Gly Leu Phe Ala Trp His Leu Asp Thr Pro Val Val Arg

```
    610                 615                 620
Ser Ala Gly Gly Arg Leu Cys Phe Leu Met Leu Gly Ser Leu Ala Ala
625                 630                 635                 640
Gly Ser Gly Ser Leu Tyr Gly Phe Phe Gly Glu Pro Thr Arg Pro Ala
                645                 650                 655
Cys Leu Leu Arg Gln Ala Leu Phe Ala Leu Gly Phe Thr Ile Phe Leu
            660                 665                 670
Ser Cys Leu Thr Val Arg Ser Phe Gln Leu Ile Ile Ile Phe Lys Phe
        675                 680                 685
Ser Thr Lys Val Pro Thr Phe Tyr His Ala Trp Val Gln Asn His Gly
    690                 695                 700
Ala Gly Leu Phe Val Met Ile Ser Ser Ala Ala Gln Leu Leu Ile Cys
705                 710                 715                 720
Leu Thr Trp Leu Val Val Trp Thr Pro Leu Pro Ala Arg Glu Tyr Gln
                725                 730                 735
Arg Phe Pro His Leu Val Met Leu Glu Cys Thr Glu Thr Asn Ser Leu
            740                 745                 750
Gly Phe Ile Leu Ala Phe Leu Tyr Asn Gly Leu Leu Ser Ile Ser Ala
        755                 760                 765
Phe Ala Cys Ser Tyr Leu Gly Lys Asp Leu Pro Glu Asn Tyr Asn Glu
    770                 775                 780
Ala Lys Cys Val Thr Phe Ser Leu Leu Phe Asn Phe Val Ser Trp Ile
785                 790                 795                 800
Ala Phe Phe Thr Thr Ala Ser Val Tyr Asp Gly Lys Tyr Leu Pro Ala
                805                 810                 815
Ala Asn Met Met Ala Gly Leu Ser Ser Leu Ser Ser Gly Phe Gly Gly
            820                 825                 830
Tyr Phe Leu Pro Lys Cys Tyr Val Ile Leu Cys Arg Pro Asp Leu Asn
        835                 840                 845
Ser Thr Glu His Phe Gln Ala Ser Ile Gln Asp Tyr Thr Arg Arg Cys
    850                 855                 860
Gly Ser Thr
865
```

<210> 32
<211> 866
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence; note =
synthetic construct

<400> 32

```
Met Val Arg Leu Leu Leu Ile Phe Phe Pro Met Ile Phe Leu Glu Met
 1               5                  10                  15
Ser Ile Leu Pro Arg Met Pro Asp Arg Lys Val Leu Leu Ala Gly Ala
            20                  25                  30
Ser Ser Gln Arg Ser Val Ala Arg Met Asp Gly Asp Val Ile Ile Gly
        35                  40                  45
Ala Leu Phe Ser Val His His Gln Pro Pro Ala Glu Lys Val Pro Glu
    50                  55                  60
Arg Lys Cys Gly Glu Ile Arg Glu Gln Tyr Gly Ile Gln Arg Val Glu
65                  70                  75                  80
Ala Met Phe His Thr Leu Asp Lys Ile Asn Ala Asp Pro Val Leu Leu
                85                  90                  95
Pro Asn Ile Thr Leu Gly Ser Glu Ile Arg Asp Ser Cys Trp His Ser
```

```
                100                     105                     110
Ser Val Ala Leu Glu Gln Ser Ile Glu Phe Ile Arg Asp Ser Leu Ile
            115                     120                     125
Ser Ile Arg Asp Glu Lys Asp Gly Leu Asn Arg Cys Leu Pro Asp Gly
        130                     135                     140
Gln Thr Leu Pro Pro Gly Arg Thr Lys Lys Pro Ile Ala Gly Val Ile
145                     150                     155                 160
Gly Pro Gly Ser Ser Ser Val Ala Ile Gln Val Gln Asn Leu Leu Gln
                    165                     170                     175
Leu Phe Asp Ile Pro Gln Ile Ala Tyr Ser Ala Thr Ser Ile Asp Leu
                180                     185                     190
Ser Asp Lys Thr Leu Tyr Lys Tyr Phe Leu Arg Val Val Pro Ser Asp
            195                     200                     205
Thr Leu Gln Ala Arg Ala Met Leu Asp Ile Val Lys Arg Tyr Asn Trp
        210                     215                     220
Thr Tyr Val Ser Ala Val His Thr Glu Gly Asn Tyr Gly Glu Ser Gly
225                     230                     235                 240
Met Asp Ala Phe Lys Glu Leu Ala Ala Gln Glu Gly Leu Cys Ile Ala
                    245                     250                     255
His Ser Asp Lys Ile Tyr Ser Asn Ala Gly Glu Lys Ser Phe Asp Arg
                260                     265                     270
Leu Leu Arg Lys Leu Arg Glu Arg Leu Pro Lys Ala Arg Val Val Val
            275                     280                     285
Cys Phe Cys Glu Gly Met Thr Val Arg Gly Leu Leu Ser Ala Met Arg
        290                     295                     300
Arg Leu Gly Val Val Gly Glu Phe Ser Leu Ile Gly Ser Asp Gly Trp
305                     310                     315                 320
Ala Asp Arg Asp Glu Val Ile Glu Gly Tyr Glu Val Glu Ala Asn Gly
                    325                     330                     335
Gly Ile Thr Ile Lys Leu Gln Ser Pro Glu Val Arg Ser Phe Asp Asp
                340                     345                     350
Tyr Phe Leu Lys Leu Arg Leu Asp Thr Asn Thr Arg Asn Pro Trp Phe
            355                     360                     365
Pro Glu Phe Trp Gln His Arg Phe Gln Cys Arg Leu Pro Gly His Leu
        370                     375                     380
Leu Glu Asn Pro Asn Phe Lys Lys Val Cys Thr Gly Asn Glu Ser Leu
385                     390                     395                 400
Glu Glu Asn Tyr Val Gln Asp Ser Lys Met Gly Phe Val Ile Asn Ala
                    405                     410                     415
Ile Tyr Ala Met Ala His Gly Leu Gln Asn Met His His Ala Leu Cys
                420                     425                     430
Pro Gly His Val Gly Leu Cys Asp Ala Met Lys Pro Ile Asp Gly Arg
            435                     440                     445
Lys Leu Leu Asp Phe Leu Ile Lys Ser Ser Phe Val Gly Val Ser Gly
        450                     455                     460
Glu Glu Val Trp Phe Asp Glu Lys Gly Asp Ala Pro Gly Arg Tyr Asp
465                     470                     475                 480
Ile Met Asn Leu Gln Tyr Thr Glu Ala Asn Arg Tyr Asp Tyr Val His
                    485                     490                     495
Val Gly Thr Trp His Glu Gly Val Leu Asn Ile Asp Asp Tyr Lys Ile
                500                     505                     510
Gln Met Asn Lys Ser Gly Met Val Arg Ser Val Cys Ser Glu Pro Cys
            515                     520                     525
Leu Lys Gly Gln Ile Lys Val Ile Arg Lys Gly Glu Val Ser Cys Cys
        530                     535                     540
Trp Ile Cys Thr Ala Cys Lys Glu Asn Glu Phe Val Gln Asp Glu Phe
545                     550                     555                 560
```

Thr Cys Arg Ala Cys Asp Leu Gly Trp Trp Pro Asn Ala Glu Leu Thr
565 570 575
Gly Cys Glu Pro Ile Pro Val Arg Tyr Leu Glu Trp His Glu Ala Pro
580 585 590
Thr Ile Ala Val Ala Leu Leu Ala Ala Leu Gly Phe Leu Ser Thr Leu
595 600 605
Ala Ile Leu Val Ile Phe Trp Arg His Phe Gln Thr Pro Ile Val Arg
610 615 620
Ser Ala Gly Gly Pro Met Cys Phe Leu Met Leu Thr Leu Leu Leu Val
625 630 635 640
Ala Tyr Met Val Val Pro Val Tyr Val Gly Pro Pro Lys Val Ser Thr
645 650 655
Cys Leu Cys Arg Gln Ala Leu Phe Pro Leu Cys Phe Thr Ile Cys Ile
660 665 670
Ser Cys Ile Ala Val Arg Ser Phe Gln Ile Val Cys Ala Phe Lys Met
675 680 685
Ala Ser Arg Phe Pro Arg Ala Tyr Ser Tyr Trp Val Arg Tyr Gln Gly
690 695 700
Pro Tyr Val Ser Met Ala Phe Ile Thr Val Leu Lys Met Val Ile Val
705 710 715 720
Val Ile Gly Met Leu Ala Thr Gly Leu Ser Pro Thr Thr Arg Thr Asp
725 730 735
Pro Asp Asp Pro Lys Ile Thr Ile Val Ser Cys Asn Pro Asn Tyr Arg
740 745 750
Asn Ser Leu Leu Phe Asn Thr Ser Leu Asp Leu Leu Leu Ser Val Val
755 760 765
Gly Phe Ser Phe Ala Tyr Met Gly Lys Glu Leu Pro Thr Asn Tyr Asn
770 775 780
Glu Ala Lys Phe Ile Thr Leu Ser Met Thr Phe Tyr Phe Thr Ser Ser
785 790 795 800
Val Ser Leu Cys Thr Phe Met Ser Ala Tyr Ser Gly Val Leu Val Thr
805 810 815
Ile Val Asp Leu Leu Val Thr Val Leu Asn Leu Leu Ala Ile Ser Leu
820 825 830
Gly Tyr Phe Gly Pro Lys Cys Tyr Met Ile Leu Phe Tyr Pro Glu Arg
835 840 845
Asn Thr Pro Ala Tyr Phe Asn Ser Met Ile Gln Gly Tyr Thr Met Arg
850 855 860
Arg Asp
865

<210> 33
<211> 876
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence; note =
synthetic construct

<400> 33
Met Val Arg Leu Leu Leu Ile Phe Phe Pro Met Ile Phe Leu Glu Met
1 5 10 15
Ser Ile Leu Pro Arg Met Pro Asp Arg Lys Val Leu Leu Ala Gly Ala
20 25 30
Ser Ser Gln Arg Ser Val Ala Arg Met Asp Gly Asp Val Ile Ile Gly
35 40 45

Ala Leu Phe Ser Val His His Gln Pro Pro Ala Glu Lys Val Pro Glu
50 55 60
Arg Lys Cys Gly Glu Ile Arg Glu Gln Tyr Gly Ile Gln Arg Val Glu
65 70 75 80
Ala Met Phe His Thr Leu Asp Lys Ile Asn Ala Asp Pro Val Leu Leu
85 90 95
Pro Asn Ile Thr Leu Gly Ser Glu Ile Arg Asp Ser Cys Trp His Ser
100 105 110
Ser Val Ala Leu Glu Gln Ser Ile Glu Phe Ile Arg Asp Ser Leu Ile
115 120 125
Ser Ile Arg Asp Glu Lys Asp Gly Leu Asn Arg Cys Leu Pro Asp Gly
130 135 140
Gln Thr Leu Pro Pro Gly Arg Thr Lys Lys Pro Ile Ala Gly Val Ile
145 150 155 160
Gly Pro Gly Ser Ser Ser Val Ala Ile Gln Val Gln Asn Leu Leu Gln
165 170 175
Leu Phe Asp Ile Pro Gln Ile Ala Tyr Ser Ala Thr Ser Ile Asp Leu
180 185 190
Ser Asp Lys Thr Leu Tyr Lys Tyr Phe Leu Arg Val Val Pro Ser Asp
195 200 205
Thr Leu Gln Ala Arg Ala Met Leu Asp Ile Val Lys Arg Tyr Asn Trp
210 215 220
Thr Tyr Val Ser Ala Val His Thr Glu Gly Asn Tyr Gly Glu Ser Gly
225 230 235 240
Met Asp Ala Phe Lys Glu Leu Ala Ala Gln Glu Gly Leu Cys Ile Ala
245 250 255
His Ser Asp Lys Ile Tyr Ser Asn Ala Gly Glu Lys Ser Phe Asp Arg
260 265 270
Leu Leu Arg Lys Leu Arg Glu Arg Leu Pro Lys Ala Arg Val Val Val
275 280 285
Cys Phe Cys Glu Gly Met Thr Val Arg Gly Leu Leu Ser Ala Met Arg
290 295 300
Arg Leu Gly Val Val Gly Glu Phe Ser Leu Ile Gly Ser Asp Gly Trp
305 310 315 320
Ala Asp Arg Asp Glu Val Ile Glu Gly Tyr Glu Val Glu Ala Asn Gly
325 330 335
Gly Ile Thr Ile Lys Leu Gln Ser Pro Glu Val Arg Ser Phe Asp Asp
340 345 350
Tyr Phe Leu Lys Leu Arg Leu Asp Thr Asn Thr Arg Asn Pro Trp Phe
355 360 365
Pro Glu Phe Trp Gln His Arg Phe Gln Cys Arg Leu Pro Gly His Leu
370 375 380
Leu Glu Asn Pro Asn Phe Lys Lys Val Cys Thr Gly Asn Glu Ser Leu
385 390 395 400
Glu Glu Asn Tyr Val Gln Asp Ser Lys Met Gly Phe Val Ile Asn Ala
405 410 415
Ile Tyr Ala Met Ala His Gly Leu Gln Asn Met His His Ala Leu Cys
420 425 430
Pro Gly His Val Gly Leu Cys Asp Ala Met Lys Pro Ile Asp Gly Arg
435 440 445
Lys Leu Leu Asp Phe Leu Ile Lys Ser Ser Phe Val Gly Val Ser Gly
450 455 460
Glu Glu Val Trp Phe Asp Glu Lys Gly Asp Ala Pro Gly Arg Tyr Asp
465 470 475 480
Ile Met Asn Leu Gln Tyr Thr Glu Ala Asn Arg Tyr Asp Tyr Val His
485 490 495
Val Gly Thr Trp His Glu Gly Val Leu Asn Ile Asp Asp Tyr Lys Ile

```
            500                 505                 510
Gln Met Asn Lys Ser Gly Met Val Arg Ser Val Cys Ser Glu Pro Cys
        515                 520                 525
Leu Lys Gly Gln Ile Lys Val Ile Arg Lys Gly Glu Val Ser Cys Cys
    530                 535                 540
Trp Ile Cys Thr Ala Cys Lys Glu Asn Glu Phe Val Gln Asp Glu Phe
545                 550                 555                 560
Thr Cys Arg Ala Cys Asp Leu Gly Trp Trp Pro Asn Ala Glu Leu Thr
                565                 570                 575
Gly Cys Glu Pro Ile Pro Val Arg Tyr Leu Glu Trp Gly Glu Pro Ala
            580                 585                 590
Val Leu Leu Leu Leu Leu Leu Leu Ser Leu Ala Leu Gly Leu Val Leu
        595                 600                 605
Ala Ala Leu Gly Leu Phe Val His His Arg Asp Ser Pro Leu Val Gln
    610                 615                 620
Ala Ser Gly Gly Pro Leu Ala Cys Phe Gly Leu Val Cys Leu Gly Leu
625                 630                 635                 640
Val Cys Leu Ser Val Leu Leu Phe Pro Gly Gln Pro Ser Pro Ala Arg
                645                 650                 655
Cys Leu Ala Gln Gln Pro Leu Ser His Leu Pro Leu Thr Gly Cys Leu
            660                 665                 670
Ser Thr Leu Phe Leu Gln Ala Ala Glu Ile Phe Val Glu Ser Glu Leu
        675                 680                 685
Pro Leu Ser Trp Ala Asp Arg Leu Ser Gly Cys Leu Arg Gly Pro Trp
    690                 695                 700
Ala Trp Leu Val Val Leu Leu Ala Met Leu Val Glu Val Ala Leu Cys
705                 710                 715                 720
Thr Trp Tyr Leu Val Ala Phe Pro Pro Glu Val Val Thr Asp Trp His
                725                 730                 735
Met Leu Pro Thr Glu Ala Leu Val His Cys Arg Thr Arg Ser Trp Val
            740                 745                 750
Ser Phe Gly Leu Ala His Ala Thr Asn Ala Thr Leu Ala Phe Leu Cys
        755                 760                 765
Phe Leu Gly Thr Phe Leu Val Arg Ser Gln Pro Gly Arg Tyr Asn Arg
    770                 775                 780
Ala Arg Gly Leu Thr Phe Ala Met Leu Ala Tyr Phe Ile Thr Trp Val
785                 790                 795                 800
Ser Phe Val Pro Leu Leu Ala Asn Val Gln Val Val Leu Arg Pro Ala
                805                 810                 815
Val Gln Met Gly Ala Leu Leu Leu Cys Val Leu Gly Ile Leu Ala Ala
            820                 825                 830
Phe His Leu Pro Arg Cys Tyr Leu Leu Met Arg Gln Pro Gly Leu Asn
        835                 840                 845
Thr Pro Glu Phe Phe Leu Gly Gly Gly Pro Gly Asp Ala Gln Gly Gln
    850                 855                 860
Asn Asp Gly Asn Thr Gly Asn Gln Gly Lys His Glu
865                 870                 875
```

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C07K 1/00, C07H 21/04, C12P 21/06

A1

(11) International Publication Number: WO 00/06592

(43) International Publication Date: 10 February 2000 (10.02.00)

(21) International Application Number: PCT/US99/17099

(22) International Filing Date: 27 July 1999 (27.07.99)

(30) Priority Data:
60/094,465 28 July 1998 (28.07.98) US

(71) Applicants (*for all designated States except US*): THE REGENTS OF THE UNIVERSITY OF CALIFORNIA [US/US]; 12th floor, 1111 Franklin Street, Oakland, CA 94607-5200 (US). THE GOVERNMENT OF THE UNITED STATES OF AMERICA, represented by THE SECRETARY, DEPARTMENT OF HEALTH AND HUMAN SERVICES [US/US]; Bethesda, MD 20892 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): ZUKER, Charles, S. [US/US]; 4778 Thurston Place, San Diego, CA 92130 (US). ADLER, Jon, Elliott [US/US]; 1099 Turquoise #10, Pacific Beach, CA 92109 (US). LINDEMEIER, Juergen [DE/DE]; Franziskaneranger 2, D-59457 Werl (DE). RYBA, Nick [US/US]; 9202 Lundigen Court, Bethesda, MD 20817 (US). HOON, Mark [US/US]; 4218 Warner Street, Kensington, MD 20895 (US).

(74) Agents: PARENT, Annette, S. et al.; Townsend and Townsend and Crew LLP, 8th floor, Two Embarcadero Center, San Francisco, CA 94107 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
With amended claims.

(54) Title: NUCLEIC ACIDS ENCODING A G-PROTEIN COUPLED RECEPTOR INVOLVED IN SENSORY TRANSDUCTION

(57) Abstract

The invention provides isolated nucleic acid and amino acid sequences of sensory cell specific G-protein coupled receptors, antibodies to such receptors, methods of detecting such nucleic acids and receptors, and methods of screening for modulators of sensory cell specific G-protein coupled receptors.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

NUCLEIC ACIDS ENCODING A G-PROTEIN COUPLED RECEPTOR INVOLVED IN SENSORY TRANSDUCTION

CROSS-REFERENCES TO RELATED APPLICATIONS

This application claims priority to USSN 60/094,465, filed July 28, 1998, herein incorporated by reference in its entirety.

STATEMENT AS TO FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

This invention was made with government support under Grant No. 5R01 DC03160, awarded by the National Institutes of Health. The government has certain rights in this invention.

FIELD OF THE INVENTION

The invention provides isolated nucleic acid and amino acid sequences of sensory cell specific G-protein coupled receptors, antibodies to such receptors, methods of detecting such nucleic acids and receptors, and methods of screening for modulators of sensory cell specific G-protein coupled receptors.

BACKGROUND OF THE INVENTION

Taste transduction is one of the most sophisticated forms of chemotransduction in animals (*see, e.g.*, Margolskee, *BioEssays* 15:645-650 (1993); Avenet & Lindemann, *J. Membrane Biol.* 112:1-8 (1989)). Gustatory signaling is found throughout the animal kingdom, from simple metazoans to the most complex of vertebrates; its main purpose is to provide a reliable signaling response to non-volatile ligands. Each of these modalities is though to be mediated by distinct signaling pathways mediated by receptors or channels, leading to receptor cell depolarization, generation of a receptor or action potential, and release of neurotransmitter at gustatory afferent neuron synapses (*see, e.g.*, Roper, *Ann. Rev. Neurosci.* 12:329-353 (1989)).

Mammals are believed to have five basic taste modalities: sweet, bitter, sour, salty and unami (the taste of monosodium glutamate) (*see, e.g.*, Kawamura & Kare, *Introduction to Unami: A Basic Taste* (1987); Kinnamon & Cummings, *Ann. Rev. Physiol.* 54:715-731(1992); Lindemann, *Physiol. Rev.* 76:718-766 (1996); Stewart *et al.*, *Am. J. Physiol.* 272:1-26 (1997)). Extensive psychophysical studies in humans have reported that different regions of the tongue display different gustatory preferences (*see, e.g.*, Hoffmann, *Menchen. Arch. Path. Anat. Physiol.* 62:516-530 (1875); Bradley *et al.*, *Anatomical Record* 212: 246-249 (1985); Miller & Reedy, *Physiol. Behav.* 47:1213-1219 (1990)). Also, numerous physiological studies in animals have shown that taste receptor cells may selectively respond to different tastants (*see, e.g.*, Akabas *et al.*, *Science* 242:1047-1050 (1988); Gilbertson *et al.*, *J. Gen. Physiol.* 100:803-24 (1992); Bernhardt *et al.*, *J. Physiol.* 490:325-336 (1996); Cummings *et al.*, *J. Neurophysiol.* 75:1256-1263 (1996)).

In mammals, taste receptor cells are assembled into taste buds that are distributed into different papillae in the tongue epithelium. Circumvallate papillae, found at the very back of the tongue, contain hundreds (mice) to thousands (human) of taste buds and are particularly sensitive to bitter substances. Foliate papillae, localized to the posterior lateral edge of the tongue, contain dozens to hundreds of taste buds and are particularly sensitive to sour and bitter substances. Fungiform papillae containing a single or a few taste buds are at the front of the tongue and are thought to mediate much of the sweet taste modality.

Each taste bud, depending on the species, contain 50-150 cells, including precursor cells, support cells, and taste receptor cells (*see, e.g.*, Lindemann, *Physiol. Rev.* 76:718-766 (1996)). Receptor cells are innervated at their base by afferent nerve endings that transmit information to the taste centers of the cortex through synapses in the brain stem and thalamus. Elucidating the mechanisms of taste cell signaling and information processing is critical for understanding the function, regulation, and "perception" of the sense of taste.

Although much is known about the psychophysics and physiology of taste cell function, very little is known about the molecules and pathways that mediate these sensory signaling responses (reviewed by Gilbertson, *Current Opn. in Neurobiol.* 3:532-539 (1993)). Electrophysiological studies suggest that sour and salty tastants modulate taste cell function by direct entry of H^+ and Na^+ ions through specialized membrane channels on the apical surface of the cell. In the case of sour compounds, taste cell

depolarization is hypothesized to result from H^+ blockage of K^+ channels (*see, e.g.*, Kinnamon *et al.*, *Proc. Nat'l Acad. Sci. USA* 85: 7023-7027 (1988)) or activation of pH-sensitive channels (*see, e.g.*, Gilbertson *et al.*, *J. Gen. Physiol.* 100:803-24 (1992)); salt transduction may be partly mediated by the entry of Na^+ via amiloride-sensitive Na^+ channels (*see, e.g.*, Heck *et al.*, *Science* 223:403-405 (1984); Brand *et al.*, *Brain Res.* 207-214 (1985); Avenet *et al.*, *Nature* 331: 351-354 (1988)).

Sweet, bitter, and unami transduction are believed to be mediated by G-protein-coupled receptor (GPCR) signaling pathways (*see, e.g.*, Striem *et al.*, *Biochem. J.* 260:121-126 (1989); Chaudhari *et al.*, *J. Neuros.* 16:3817-3826 (1996); Wong *et al.*, *Nature* 381: 796-800 (1996)). Confusingly, there are almost as many models of signaling pathways for sweet and bitter transduction as there are effector enzymes for GPCR cascades (e.g., G protein subunits, cGMP phosphodiesterase, phospholipase C, adenylate cyclase; *see, e.g.*, Kinnamon & Margolskee, *Curr. Opin. Neurobiol.* 6:506-513 (1996)). However, little is known about the specific membrane receptors involved in taste transduction, or many of the individual intracellular signaling molecules activated by the individual taste transduction pathways. Identification of such molecules is important given the numerous pharmacological and food industry applications for bitter antagonists, sweet agonists, and modulators of salty and sour taste.

The identification and isolation of taste receptors (including taste ion channels), and taste signaling molecules, such as G-protein subunits and enzymes involved in signal transduction, would allow for the pharmacological and genetic modulation of taste transduction pathways. For example, availability of receptor and channel molecules would permit the screening for high affinity agonists, antagonists, inverse agonists, and modulators of taste cell activity. Such taste modulating compounds could then be used in the pharmaceutical and food industries to customize taste. In addition, such taste cell specific molecules can serve as invaluable tools in the generation of taste topographic maps that elucidate the relationship between the taste cells of the tongue and taste sensory neurons leading to taste centers in the brain.

SUMMARY OF THE INVENTION

The present invention thus provides for the first time nucleic acids encoding a taste cell specific G-protein coupled receptor. These nucleic acids and the polypeptides that they encode are referred to as "GPCR-B3" for G-protein coupled

receptor ("GPCR") B3. These taste cell specific GPCRs are components of the taste transduction pathway.

In one aspect, the present invention provides an isolated nucleic acid encoding a sensory transduction G-protein coupled receptor, the receptor comprising greater than about 70% amino acid identity to an amino acid sequence of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3.

In one embodiment, the nucleic acid comprises a nucleotide sequence of SEQ ID NO:4, SEQ ID NO:5, or SEQ ID NO:6. In another embodiment, the nucleic acid is amplified by primers that selectively hybridize under stringent hybridization conditions to the same sequence as degenerate primer sets encoding amino acid sequences selected from the group consisting of: IAWDWNGPKW (SEQ ID NO:7) and LPENYNEAKC (SEQ ID NO:8).

In another aspect, the present invention provides an isolated nucleic acid encoding a sensory transduction G-protein coupled receptor, wherein the nucleic acid specifically hybridizes under highly stringent conditions to a nucleic acid having the sequence of SEQ ID NO:4, SEQ ID NO:5, or SEQ ID NO:6.

In another aspect, the present invention provides an isolated nucleic acid encoding a sensory transduction G-protein coupled receptor, the receptor comprising greater than about 70% amino acid identity to a polypeptide having a sequence of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3 wherein the nucleic acid selectively hybridizes under moderately stringent hybridization conditions to a nucleotide sequence of SEQ ID NO:4, SEQ ID NO:5, or SEQ ID NO:6.

In another aspect, the present invention provides an isolated nucleic acid encoding an extracellular domain of a sensory transduction G-protein coupled receptor, the extracellular domain having greater than about 70% amino acid sequence identity to the extracellular domain of SEQ ID NO:1.

In another aspect, the present invention provides an isolated nucleic acid encoding a transmembrane domain of a sensory transduction G-protein coupled receptor, the transmembrane domain comprising greater than about 70% amino acid sequence identity to the transmembrane domain of SEQ ID NO:1.

In another aspect, the present invention provides an isolated sensory transduction G-protein coupled receptor, the receptor comprising greater than about 70%

amino acid sequence identity to an amino acid sequence of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3.

In one embodiment, the receptor specifically binds to polyclonal antibodies generated against SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3. In another embodiment, the receptor has G-protein coupled receptor activity. In another embodiment, the receptor has an amino acid sequence of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3. In another embodiment, the receptor is from a human, a rat, or a mouse.

In one aspect, the present invention provides an isolated polypeptide comprising an extracellular domain of a sensory transduction G-protein coupled receptor, the extracellular domain comprising greater than about 70% amino acid sequence identity to the extracellular domain of SEQ ID NO:1.

In one embodiment, the polypeptide encodes the extracellular domain of SEQ ID NO:1. In another embodiment, the extracellular domain is covalently linked to a heterologous polypeptide, forming a chimeric polypeptide.

In one aspect, the present invention provides an isolated polypeptide comprising a transmembrane domain of a sensory transduction G-protein coupled receptor, the transmembrane domain comprising greater than about 70% amino acid sequence identity to the transmembrane domain of SEQ ID NO:1.

In one embodiment, the polypeptide encodes the transmembrane domain of SEQ ID NO:1. In another embodiment, the polypeptide further comprises a cytoplasmic domain comprising greater than about 70% amino acid identity to the cytoplasmic domain of SEQ ID NO:1. In another embodiment, the polypeptide encodes the cytoplasmic domain of SEQ ID NO:1. In another embodiment, the transmembrane domain is covalently linked to a heterologous polypeptide, forming a chimeric polypeptide. In another embodiment, the chimeric polypeptide has G-protein coupled receptor activity.

In one aspect, the present invention provides an antibody that selectively binds to the receptor comprising greater than about 70% amino acid sequence identity to an amino acid sequence of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3.

In another aspect, the present invention provides an expression vector comprising a nucleic acid encoding a polypeptide comprising greater than about 70% amino acid sequence identity to an amino acid sequence of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3.

In another aspect, the present invention provides a host cell transfected with the expression vector.

In another aspect, the present invention provides a method for identifying a compound that modulates sensory signaling in sensory cells, the method comprising the steps of: (i) contacting the compound with a polypeptide comprising an extracellular domain of a sensory transduction G-protein coupled receptor, the extracellular domain comprising greater than about 70% amino acid sequence identity to the extracellular domain of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3; and (ii) determining the functional effect of the compound upon the extracellular domain.

In another aspect, the present invention provides a method for identifying a compound that modulates sensory signaling in sensory cells, the method comprising the steps of: (i) contacting the compound with a polypeptide comprising an extracellular domain of a sensory transduction G-protein coupled receptor, the transmembrane domain comprising greater than about 70% amino acid sequence identity to the extracellular domain of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3; and (ii) determining the functional effect of the compound upon the transmembrane domain.

In one embodiment, the polypeptide is a sensory transduction G-protein coupled receptor, the receptor comprising greater than about 70% amino acid identity to a polypeptide encoding SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3. In another embodiment, polypeptide comprises an extracellular domain that is covalently linked to a heterologous polypeptide, forming a chimeric polypeptide. In another embodiment, the polypeptide has G-protein coupled receptor activity. In another embodiment, the extracellular domain is linked to a solid phase, either covalently or non-covalently. In another embodiment, the functional effect is determined by measuring changes in intracellular cAMP, IP3, or Ca^{2+}. In another embodiment, the functional effect is a chemical effect. In another embodiment, the functional effect is a chemical effect. In another embodiment, the functional effect is determined by measuring binding of the compound to the extracellular domain. In another embodiment, the polypeptide is recombinant. In another embodiment, the polypeptide is expressed in a cell or cell membrane. In another embodiment, the cell is a eukaryotic cell.

In one embodiment, the polypeptide comprises an transmembrane domain that is covalently linked to a heterologous polypeptide, forming a chimeric polypeptide.

In one aspect, the present invention provides a method of making a sensory transduction G-protein coupled receptor, the method comprising the step of

expressing the receptor from a recombinant expression vector comprising a nucleic acid encoding the receptor, wherein the amino acid sequence of the receptor comprises greater than about 70% amino acid identity to a polypeptide having a sequence of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3.

In one aspect, the present invention provides a method of making a recombinant cell comprising a sensory transduction G-protein coupled receptor, the method comprising the step of transducing the cell with an expression vector comprising a nucleic acid encoding the receptor, wherein the amino acid sequence of the receptor comprises greater than about 70% amino acid identity to a polypeptide having a sequence of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3.

In one aspect, the present invention provides a method of making an recombinant expression vector comprising a nucleic acid encoding a sensory transduction G-protein coupled receptor, the method comprising the step of ligating to an expression vector a nucleic acid encoding the receptor, wherein the amino acid sequence of the receptor comprises greater than about 70% amino acid identity to a polypeptide having a sequence of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the proposed topology of GPCR-B3, with a large extracellular domain extending from amino acid 1 to about amino acid 580 of the rat GPCR-B3 amino acid sequence (corresponding to nucleotide residues 1-1740 of the rat sequence, with the ATG initiator methionine defined as residue 1), and seven transmembrane domains. The large extracellular domain may extend into the first transmembrane domain. Dark residues indicate identities between GPCR-B3 and GPCR-B4 (for a description of GPCR-B4, *see, e.g.*, USSN 60/095,464, filed July 28, 1998, and USSN 60/112,747, filed December 17, 1998; *see also* Hoon *et al., Cell* 96:541-551 (1999)).

Figure 2 is a western blot showing GPCR-B3 protein expression in taste buds but not in non-taste tissue. Using PCR assays, the following non-tongue tissues were screened for GPCR-B3 expression--brain, liver, olfactory epithelium, VNO, and heart. GPCR-B3 was expressed only in taste tissue (data not shown).

Figure 3 shows *in situ* hybridization of tongue tissue sections showing labeling of GPCR-B3 in taste receptor cells of taste buds, but not in adjacent non-taste tissue.

Figure 4 shows a chimeric receptor containing the entire extracellular domain of the murine mGluR1 receptor and the transmembrane domain comprising seven transmembrane regions and corresponding cytosolic loops, and C-terminal end from murine GPCR-B3.

Figure 5 shows HEK cells transfected with the chimeric glutamate/GPCR-B3 receptor described in Figure 4. Figure 5 shows calcium response to glutamate, demonstrating robust coupling of the chimeric receptor to phospholipase C. These results indicate that the chimeric glutamate/GPCR-B3 can couple to the promiscuous G protein $G\alpha 15$ and trigger calcium responses that are detectable using the indicator Fura-2.

DETAILED DESCRIPTION OF THE INVENTION

I. Introduction

The present invention provides for the first time nucleic acids encoding a taste cell specific G-protein coupled receptor. These nucleic acids and the receptors that they encode are referred to as "GPCR" for G-protein coupled receptor, and are designated as GPCR-B3. These taste cell specific GPCR are components of the taste transduction pathway. These nucleic acids provide valuable probes for the identification of taste cells, as the nucleic acids are specifically expressed in taste cells. For example, probes for GPCR polypeptides and proteins can be used to identity subsets of taste cells such as foliate cells and circumvallate cells, or specific taste receptor cells, e.g., sweet, sour, salty, and bitter. They also serve as tools for the generation of taste topographic maps that elucidate the relationship between the taste cells of the tongue and taste sensory neurons leading to taste centers in the brain. Furthermore, the nucleic acids and the proteins they encode can be used as probes to dissect taste-induced behaviors.

The invention also provides methods of screening for modulators, e.g., activators, inhibitors, stimulators, enhancers, agonists, and antagonists, of these novel taste cell GPCRs. Such modulators of taste transduction are useful for pharmacological and genetic modulation of taste signaling pathways. These methods of screening can be used to identify high affinity agonists and antagonists of taste cell activity. These modulatory compounds can then be used in the food and pharmaceutical industries to

customize taste. Thus, the invention provides assays for taste modulation, where GPCR-B3 acts as an direct or indirect reporter molecule for the effect of modulators on taste transduction. GPCRs can be used in assays, e.g., to measure changes in ion concentration, membrane potential, current flow, ion flux, transcription, signal transduction, receptor-ligand interactions, second messenger concentrations, *in vitro*, *in vivo*, and *ex vivo*. In one embodiment, GPCR-B3 can be used as an indirect reporter via attachment to a second reporter molecule such as green fluorescent protein (*see, e.g.*, Mistili & Spector, *Nature Biotechnology* 15:961-964 (1997)). In another embodiment, GPCR-B3 is recombinantly expressed in cells, and modulation of taste transduction via GPCR activity is assayed by measuring changes in Ca^{2+} levels.

Methods of assaying for modulators of taste transduction include *in vitro* ligand binding assays using GPCR-B3, portions thereof such as the extracellular domain, or chimeric proteins comprising one or more domains of GPCR-B3, oocyte GPCR-B3 expression; tissue culture cell GPCR-B3 expression; transcriptional activation of GPCR-B3; phosphorylation and dephosphorylation of GPCRs; G-protein binding to GPCRs; ligand binding assays; voltage, membrane potential and conductance changes; ion flux assays; changes in intracellular second messengers such as cAMP and inositol triphosphate; changes in intracellular calcium levels; and neurotransmitter release.

Finally, the invention provides for methods of detecting GPCR-B3 nucleic acid and protein expression, allowing investigation of taste transduction regulation and specific identification of taste receptor cells. GPCR-B3 also provides useful nucleic acid probes for paternity and forensic investigations. GPCR-B3 is a useful nucleic acid probe for identifying subpopulations of taste receptor cells such as foliate, fungiform, and circumvallate taste receptor cells. GPCR-B3 receptors can also be used to generate monoclonal and polyclonal antibodies useful for identifying taste receptor cells. Taste receptor cells can be identified using techniques such as reverse transcription and amplification of mRNA, isolation of total RNA or poly A^+ RNA, northern blotting, dot blotting, *in situ* hybridization, RNase protection, S1 digestion, probing DNA microchip arrays, western blots, and the like.

Functionally, GPCR-B3 represents a seven transmembrane G-protein coupled receptor involved in taste transduction, which interacts with a G-protein to mediate taste signal transduction (*see, e.g.*, Fong, *Cell Signal* 8:217 (1996); Baldwin, *Curr. Opin. Cell Biol.* 6:180 (1994)).

Structurally, the nucleotide sequence of GPCR-B3 (see, e.g., SEQ ID NOS:4-6, isolated from rat, mouse, and human respectively) encodes a polypeptide of approximately 840 amino acids with a predicted molecular weight of approximately 97 kDa and a predicted range of 92-102 kDa (see, e.g., SEQ ID NOS:1-3). Related GPCR-B3 genes from other species share at least about 70% amino acid identity over a amino acid region at least about 25 amino acids in length, optionally 50 to 100 amino acids in length. GPCR-B3 is specifically expressed in foliate and fungiform cells, with lower expression in circumvallate taste receptor cells of the tongue. GPCR-B3 is an moderately rare sequence found in approximately 1/150,000 cDNAs from an oligo-dT primed circumvallate cDNA library (see Example 1).

The present invention also provides polymorphic variants of the GPCR-B3 depicted in SEQ ID NO:1: variant #1, in which an isoleucine residue is substituted for a leucine acid residue at amino acid position 33; variant #2, in which an aspartic acid residue is substituted for a glutamic acid residue at amino acid position 84; and variant #3, in which a glycine residue is substituted for an alanine residue at amino acid position 90.

Specific regions of the GPCR-B3 nucleotide and amino acid sequence may be used to identify polymorphic variants, interspecies homologs, and alleles of GPCR-B3. This identification can be made *in vitro*, e.g., under stringent hybridization conditions or PCR (using primers encoding SEQ ID NOS:7-8) and sequencing, or by using the sequence information in a computer system for comparison with other nucleotide sequences. Typically, identification of polymorphic variants and alleles of GPCR-B3 is made by comparing an amino acid sequence of about 25 amino acids or more, e.g., 50-100 amino acids. Amino acid identity of approximately at least 70% or above, optionally 80%, optionally 90-95% or above typically demonstrates that a protein is a polymorphic variant, interspecies homolog, or allele of GPCR-B3. Sequence comparison can be performed using any of the sequence comparison algorithms discussed below. Antibodies that bind specifically to GPCR-B3 or a conserved region thereof can also be used to identify alleles, interspecies homologs, and polymorphic variants.

Polymorphic variants, interspecies homologs, and alleles of GPCR-B3 are confirmed by examining taste cell specific expression of the putative GPCR-B3 polypeptide. Typically, GPCR-B3 having the amino acid sequence of SEQ ID NO:1-3 is used as a positive control in comparison to the putative GPCR-B3 protein to demonstrate the identification of a polymorphic variant or allele of GPCR-B3. The polymorphic

variants, alleles and interspecies homologs are expected to retain the seven transmembrane structure of a G-protein coupled receptor.

GPCR-B3 nucleotide and amino acid sequence information may also be used to construct models of taste cell specific polypeptides in a computer system. These models are subsequently used to identify compounds that can activate or inhibit GPCR-B3. Such compounds that modulate the activity of GPCR B4 can be used to investigate the role of GPCR-B3 in taste transduction.

The isolation of GPCR-B3 for the first time provides a means for assaying for inhibitors and activators of G-protein coupled receptor taste transduction . Biologically active GPCR-B3 is useful for testing inhibitors and activators of GPCR-B3 as taste transducers using *in vivo* and *in vitro* expression that measure, e.g., transcriptional activation of GPCR-B3; ligand binding; phosphorylation and dephosphorylation; binding to G-proteins; G-protein activation; regulatory molecule binding; voltage, membrane potential and conductance changes; ion flux; intracellular second messengers such as cAMP and inositol triphosphate; intracellular calcium levels; and neurotransmitter release. Such activators and inhibitors identified using GPCR-B3, can be used to further study taste transduction and to identify specific taste agonists and antagonists. Such activators and inhibitors are useful as pharmaceutical and food agents for customizing taste.

Methods of detecting GPCR-B3 nucleic acids and expression of GPCR-B3 are also useful for identifying taste cells and creating topological maps of the tongue and the relation of tongue taste receptor cells to taste sensory neurons in the brain. Chromosome localization of the genes encoding human GPCR-B3 can be used to identify diseases, mutations, and traits caused by and associated with GPCR-B3.

II. Definitions

As used herein, the following terms have the meanings ascribed to them unless specified otherwise.

"Taste receptor cells" are neuroepithelial cells that are organized into groups to form taste buds of the tongue, e.g., foliate, fungiform, and circumvallate cells (*see, e.g.*, Roper *et al., Ann. Rev. Neurosci.* 12:329-353 (1989)).

"GPCR-B3," also called "TR1," refers to a G-protein coupled receptor is specifically expressed in taste receptor cells such as foliate, fungiform, and circumvallate cells (*see, e.g.*, Hoon *et al., Cell* 96:541-551 (1999), herein incorporated by reference in

its entirety). Such taste cells can be identified because they express specific molecules such as Gustducin, a taste cell specific G protein (McLaughin *et al., Nature* 357:563-569 (1992)). Taste receptor cells can also be identified on the basis of morphology (*see, e.g.*, Roper, *supra*).

GPCR-B3 encodes GPCRs with seven transmembrane regions that have "G-protein coupled receptor activity," e.g., they bind to G-proteins in response to extracellular stimuli and promote production of second messengers such as IP3, cAMP, and Ca^{2+} via stimulation of enzymes such as phospholipase C and adenylate cyclase (for a description of the structure and function of GPCRs, *see, e.g.*, Fong, *supra*, and Baldwin, *supra*).

The term GPCR-B3 therefore refers to polymorphic variants, alleles, mutants, and interspecies homologs that: (1) have about 70% amino acid sequence identity, optionally about 75, 80, 85, 90, or 95% amino acid sequence identity to SEQ ID NOS:1-3 over a window of about 25 amino acids, optionally 50-100 amino acids; (2) bind to antibodies raised against an immunogen comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-3 and conservatively modified variants thereof; (3) specifically hybridize (with a size of at least about 500, optionally at least about 900 nucleotides) under stringent hybridization conditions to a sequence selected from the group consisting of SEQ ID NO:4-6, and conservatively modified variants thereof; or (4) are amplified by primers that specifically hybridize under stringent hybridization conditions to the same sequence as a degenerate primer sets encoding SEQ ID NOS:7-8.

Topologically, sensory GPCRs have an N-terminal "extracellular domain," a "transmembrane domain" comprising seven transmembrane regions and corresponding cytoplasmic and extracellular loops, and a C-terminal "cytoplasmic domain" (*see* Figure 1; *see also* Hoon *et al., Cell* 96:541-551 (1999); Buck & Axel, *Cell* 65:175-187 (1991)). These domains can be structurally identified using methods known to those of skill in the art, such as sequence analysis programs that identify hydrophobic and hydrophilic domains (*see, e.g.*, Kyte & Doolittle, *J. Mol. Biol.* 157:105-132 (1982)). Such domains are useful for making chimeric proteins and for *in vitro* assays of the invention.

"Extracellular domain" therefore refers to the domain of GPCR-B3 that protrudes from the cellular membrane and binds to extracellular ligand. This region starts at the N-terminus and ends approximately at the conserved glutamic acid at amino acid position 563 plus or minus approximately 20 amino acids. The region corresponding to

amino acids 1-580 of SEQ ID NO:1 (nucleotides 1-1740, with nucleotide 1 starting at the ATG initiator methionine codon; *see also* Figure 1) is one embodiment of an extracellular domain that extends slightly into the transmembrane domain. This embodiment is useful for *in vitro* ligand binding assays, both soluble and solid phase.

"Transmembrane domain," comprising seven transmembrane regions plus the corresponding cytoplasmic and extracellular loops, refers to the domain of GPCR-B3 that starts approximately at the conserved glutamic acid residue at amino acid position 563 plus or minus approximately 20 amino acids and ends approximately at the conserved tyrosine amino acid residue at position 812 plus or minus approximately 10 amino acids.

"Cytoplasmic domain" refers to the domain of GPCR-B3 that starts at the conserved tyrosine amino acid residue at position 812 plus or minus approximately 10 amino acids and continues to the C-terminus of the polypeptide.

"Biological sample" as used herein is a sample of biological tissue or fluid that contains GPCR-B3 or nucleic acid encoding GPCR-B3 protein. Such samples include, but are not limited to, tissue isolated from humans, mice, and rats, in particular, ton. Biological samples may also include sections of tissues such as frozen sections taken for histological purposes. A biological sample is typically obtained from a eukaryotic organism, such as insects, protozoa, birds, fish, reptiles, and preferably a mammal such as rat, mouse, cow, dog, guinea pig, or rabbit, and most preferably a primate such as chimpanzees or humans. Tissues include tongue tissue, isolated taste buds, and testis tissue.

"GPCR activity" refers to the ability of a GPCR to transduce a signal. Such activity can be measured in a heterologous cell, by coupling a GPCR (or a chimeric GPCR) to either a G-protein or promiscuous G-protein such as $G\alpha15$, and an enzyme such as PLC, and measuring increases in intracellular calcium using (Offermans & Simon, *J. Biol. Chem.* 270:15175-15180 (1995)). Receptor activity can be effectively measured by recording ligand-induced changes in $[Ca^{2+}]_i$ using fluorescent Ca^{2+}-indicator dyes and fluorometric imaging. Optionally, the polypeptides of the invention are involved in sensory transduction, optionally taste transduction in taste cells.

The phrase "functional effects" in the context of assays for testing compounds that modulate GPCR-B3 mediated taste transduction includes the determination of any parameter that is indirectly or directly under the influence of the receptor, e.g., functional, physical and chemical effects. It includes ligand binding,

changes in ion flux, membrane potential, current flow, transcription, G-protein binding, GPCR phosphorylation or dephosphorylation, signal transduction, receptor-ligand interactions, second messenger concentrations (e.g., cAMP, IP3, or intracellular Ca^{2+}), *in vitro*, *in vivo*, and *ex vivo* and also includes other physiologic effects such increases or decreases of neurotransmitter or hormone release.

By "determining the functional effect" is meant assays for a compound that increases or decreases a parameter that is indirectly or directly under the influence of GPCR-B3, e.g., functional, physical and chemical effects. Such functional effects can be measured by any means known to those skilled in the art, e.g., changes in spectroscopic characteristics (e.g., fluorescence, absorbance, refractive index), hydrodynamic (e.g., shape), chromatographic, or solubility properties, patch clamping, voltage-sensitive dyes, whole cell currents, radioisotope efflux, inducible markers, oocyte GPCR-B3 expression; tissue culture cell GPCR-B3 expression; transcriptional activation of GPCR-B3; ligand binding assays; voltage, membrane potential and conductance changes; ion flux assays; changes in intracellular second messengers such as cAMP and inositol triphosphate (IP3); changes in intracellular calcium levels; neurotransmitter release, and the like.

"Inhibitors," "activators," and "modulators" of GPCR-B3 are used interchangeably to refer to inhibitory, activating, or modulating molecules identified using *in vitro* and *in vivo* assays for taste transduction, e.g., ligands, agonists, antagonists, and their homologs and mimetics. Inhibitors are compounds that, e.g., bind to, partially or totally block stimulation, decrease, prevent, delay activation, inactivate, desensitize, or down regulate taste transduction, e.g., antagonists. Activators are compounds that, e.g., bind to, stimulate, increase, open, activate, facilitate, enhance activation, sensitize or up regulate taste transduction, e.g., agonists. Modulators include compounds that, e.g., alter the interaction of a receptor with: extracellular proteins that bind activators or inhibitor (e.g., ebnerin and other members of the hydrophobic carrier family); G -proteins; kinases (e.g., homologs of rhodopsin kinase and beta adrenergic receptor kinases that are involved in deactivation and desensitization of a receptor); and arrestin-like proteins, which also deactivate and desensitize receptors. Modulators include genetically modified versions of GPCR-B3, e.g., with altered activity, as well as naturally occurring and synthetic ligands, antagonists, agonists, small chemical molecules and the like. Such assays for inhibitors and activators include, e.g., expressing GPCR-B3 in cells or cell membranes, applying putative modulator compounds, and then determining the functional effects on taste transduction, as described above. Samples or assays comprising GPCR-

B3 that are treated with a potential activator, inhibitor, or modulator are compared to control samples without the inhibitor, activator, or modulator to examine the extent of inhibition. Control samples (untreated with inhibitors) are assigned a relative GPCR-B3 activity value of 100%. Inhibition of GPCR-B3 is achieved when the GPCR-B3 activity value relative to the control is about 80%, optionally 50%, 25-0%. Activation of GPCR-B3 is achieved when the GPCR-B3 activity value relative to the control is 110%, optionally 150%, 200-500%, 1000-3000% higher.

"Biologically active" GPCR-B3 refers to GPCR-B3 having GPCR activity as described above, involved in taste transduction in taste receptor cells.

The terms "isolated" "purified" or "biologically pure" refer to material that is substantially or essentially free from components which normally accompany it as found in its native state. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high performance liquid chromatography. A protein that is the predominant species present in a preparation is substantially purified. In particular, an isolated GPCR-B3 nucleic acid is separated from open reading frames that flank the GPCR-B3 gene and encode proteins other than GPCR-B3. The term "purified" denotes that a nucleic acid or protein gives rise to essentially one band in an electrophoretic gel. Particularly, it means that the nucleic acid or protein is at least 85% pure, optionally at least 95% pure, and optionally at least 99% pure.

"Nucleic acid" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. The term encompasses nucleic acids containing known nucleotide analogs or modified backbone residues or linkages, which are synthetic, naturally occurring, and non-naturally occurring, which have similar binding properties as the reference nucleic acid, and which are metabolized in a manner similar to the reference nucleotides. Examples of such analogs include, without limitation, phosphorothioates, phosphoramidates, methyl phosphonates, chiral-methyl phosphonates, 2-O-methyl ribonucleotides, peptide-nucleic acids (PNAs).

Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (e.g., degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer *et al., Nucleic Acid Res.* 19:5081 (1991);

Ohtsuka *et al.*, *J. Biol. Chem.* 260:2605-2608 (1985); Rossolini *et al.*, *Mol. Cell. Probes* 8:91-98 (1994)). The term nucleic acid is used interchangeably with gene, cDNA, mRNA, oligonucleotide, and polynucleotide.

The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-naturally occurring amino acid polymer.

The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, e.g., hydroxyproline, γ-carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, i.e., an α carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, e.g., homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (e.g., norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. Amino acid mimetics refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that functions in a manner similar to a naturally occurring amino acid.

Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

"Conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide.

Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the invention.

The following eight groups each contain amino acids that are conservative substitutions for one another:

1) Alanine (A), Glycine (G);
2) Aspartic acid (D), Glutamic acid (E);
3) Asparagine (N), Glutamine (Q);
4) Arginine (R), Lysine (K);
5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V);
6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W);
7) Serine (S), Threonine (T); and
8) Cysteine (C), Methionine (M)

(*see, e.g.*, Creighton, *Proteins* (1984)).

Macromolecular structures such as polypeptide structures can be described in terms of various levels of organization. For a general discussion of this organization, *see, e.g.*, Alberts *et al.*, *Molecular Biology of the Cell* (3rd ed., 1994) and Cantor and Schimmel, *Biophysical Chemistry Part I: The Conformation of Biological Macromolecules* (1980). "Primary structure" refers to the amino acid sequence of a particular peptide. "Secondary structure" refers to locally ordered, three dimensional structures within a polypeptide. These structures are commonly known as domains.

Domains are portions of a polypeptide that form a compact unit of the polypeptide and are typically 50 to 350 amino acids long. Typical domains are made up of sections of lesser organization such as stretches of β-sheet and α-helices. "Tertiary structure" refers to the complete three dimensional structure of a polypeptide monomer. "Quaternary structure" refers to the three dimensional structure formed by the noncovalent association of independent tertiary units.. Anisotropic terms are also known as energy terms.

A "label" or a "detectable moiety" is a composition detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include ^{32}P, fluorescent dyes, electron-dense reagents, enzymes (e.g., as commonly used in an ELISA), biotin, digoxigenin, or haptens and proteins for which ant or 7 can be made detectable, e.g., by incorporating a radiolabel into the peptide, and used to detect antibodies specifically reactive with the peptide).

A "labeled nucleic acid probe or oligonucleotide" is one that is bound, either covalently, through a linker or a chemical bond, or noncovalently, through ionic, van der Waals, electrostatic, or hydrogen bonds to a label such that the presence of the probe may be detected by detecting the presence of the label bound to the probe.

As used herein a "nucleic acid probe or oligonucleotide" is defined as a nucleic acid capable of binding to a target nucleic acid of complementary sequence through one or more types of chemical bonds, usually through complementary base pairing, usually through hydrogen bond formation. As used herein, a probe may include natural (i.e., A, G, C, or T) or modified bases (7-deazaguanosine, inosine, etc.). In addition, the bases in a probe may be joined by a linkage other than a phosphodiester bond, so long as it does not interfere with hybridization. Thus, for example, probes may be peptide nucleic acids in which the constituent bases are joined by peptide bonds rather than phosphodiester linkages. It will be understood by one of skill in the art that probes may bind target sequences lacking complete complementarity with the probe sequence depending upon the stringency of the hybridization conditions. The probes are optionally directly labeled as with isotopes, chromophores, lumiphores, chromogens, or indirectly labeled such as with biotin to which a streptavidin complex may later bind. By assaying for the presence or absence of the probe, one can detect the presence or absence of the select sequence or subsequence.

The term "recombinant" when used with reference, e.g., to a cell, or nucleic acid, protein, or vector, indicates that the cell, nucleic acid, protein or vector, has

been modified by the introduction of a heterologous nucleic acid or protein or the alteration of a native nucleic acid or protein, or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under expressed or not expressed at all.

The term "heterologous" when used with reference to portions of a nucleic acid indicates that the nucleic acid comprises two or more subsequences that are not found in the same relationship to each other in nature. For instance, the nucleic acid is typically recombinantly produced, having two or more sequences from unrelated genes arranged to make a new functional nucleic acid, e.g., a promoter from one source and a coding region from another source. Similarly, a heterologous protein indicates that the protein comprises two or more subsequences that are not found in the same relationship to each other in nature (e.g., a fusion protein).

A "promoter" is defined as an array of nucleic acid control sequences that direct transcription of a nucleic acid. As used herein, a promoter includes necessary nucleic acid sequences near the start site of transcription, such as, in the case of a polymerase II type promoter, a TATA element. A promoter also optionally includes distal enhancer or repressor elements, which can be located as much as several thousand base pairs from the start site of transcription. A "constitutive" promoter is a promoter that is active under most environmental and developmental conditions. An "inducible" promoter is a promoter that is active under environmental or developmental regulation. The term "operably linked" refers to a functional linkage between a nucleic acid expression control sequence (such as a promoter, or array of transcription factor binding sites) and a second nucleic acid sequence, wherein the expression control sequence directs transcription of the nucleic acid corresponding to the second sequence.

An "expression vector" is a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements that permit transcription of a particular nucleic acid in a host cell. The expression vector can be part of a plasmid, virus, or nucleic acid fragment. Typically, the expression vector includes a nucleic acid to be transcribed operably linked to a promoter.

The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same (i.e., 70% identity, optionally 75%, 80%, 85%, 90%, or 95% identity over a

specified region), when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. Such sequences are then said to be "substantially identical." This definition also refers to the compliment of a test sequence. Optionally, the identity exists over a region that is at least about 50 amino acids or nucleotides in length, or more preferably over a region that is 75-100 amino acids or nucleotides in length.

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

A "comparison window", as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection (*see, e.g., Current Protocols in Molecular Biology* (Ausubel *et al.*, eds. 1995 supplement)).

One example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments to show relationship and percent sequence identity. It also plots a tree or dendogram showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng & Doolittle, *J. Mol.*

Evol. 35:351-360 (1987). The method used is similar to the method described by Higgins & Sharp, *CABIOS* 5:151-153 (1989). The program can align up to 300 sequences, each of a maximum length of 5,000 nucleotides or amino acids. The multiple alignment procedure begins with the pairwise alignment of the two most similar sequences, producing a cluster of two aligned sequences. This cluster is then aligned to the next most related sequence or cluster of aligned sequences. Two clusters of sequences are aligned by a simple extension of the pairwise alignment of two individual sequences. The final alignment is achieved by a series of progressive, pairwise alignments. The program is run by designating specific sequences and their amino acid or nucleotide coordinates for regions of sequence comparison and by designating the program parameters. Using PILEUP, a reference sequence is compared to other test sequences to determine the percent sequence identity relationship using the following parameters: default gap weight (3.00), default gap length weight (0.10), and weighted end gaps. PILEUP can be obtained from the GCG sequence analysis software package, e.g., version 7.0 (Devereaux *et al.*, *Nuc. Acids Res.* 12:387-395 (1984).

Another example of algorithm that is suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul *et al., Nuc. Acids Res.* 25:3389-3402 (1977) and Altschul *et al., J. Mol. Biol.* 215:403-410 (1990), respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al., supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either

sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) or 10, M=5, N=-4 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength of 3, and expectation (E) of 10, and the BLOSUM62 scoring matrix (*see* Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.

The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.*, Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

An indication that two nucleic acid sequences or polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the antibodies raised against the polypeptide encoded by the second nucleic acid, as described below. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are substantially identical is that the two molecules or their complements hybridize to each other under stringent conditions, as described below. Yet another indication that two nucleic acid sequences are substantially identical is that the same primers can be used to amplify the sequence.

The phrase "selectively (or specifically) hybridizes to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent hybridization conditions when that sequence is present in a complex mixture (e.g., total cellular or library DNA or RNA).

The phrase "stringent hybridization conditions" refers to conditions under which a probe will hybridize to its target subsequence, typically in a complex mixture of nucleic acid, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at

higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Probes*, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally, stringent conditions are selected to be about 5-10°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength pH. The T_m is the temperature (under defined ionic strength, pH, and nucleic concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at T_m, 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (e.g., 10 to 50 nucleotides) and at least about 60°C for long probes (*e.g.*, greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. For selective or specific hybridization, a positive signal is at least two times background, optionally 10 times background hybridization. Exemplary stringent hybridization conditions can be as following: 50% formamide, 5x SSC, and 1% SDS, incubating at 42°C, or, 5x SSC, 1% SDS, incubating at 65°C, with wash in 0.2x SSC, and 0.1% SDS at 65°C.

Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, for example, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. In such cases, the nucleic acids typically hybridize under moderately stringent hybridization conditions. Exemplary "moderately stringent hybridization conditions" include a hybridization in a buffer of 40% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 1X SSC at 45°C. A positive hybridization is at least twice background. Those of ordinary skill will readily recognize that alternative hybridization and wash conditions can be utilized to provide conditions of similar stringency.

"Antibody" refers to a polypeptide comprising a framework region from an immunoglobulin gene or fragments thereof that specifically binds and recognizes an antigen. The recognized immunoglobulin genes include the kappa, lambda, alpha, gamma, delta, epsilon, and mu constant region genes, as well as the myriad immunoglobulin variable region genes. Light chains are classified as either kappa or

lambda. Heavy chains are classified as gamma, mu, alpha, delta, or epsilon, which in turn define the immunoglobulin classes, IgG, IgM, IgA, IgD and IgE, respectively.

An exemplary immunoglobulin (antibody) structural unit comprises a tetramer. Each tetramer is composed of two identical pairs of polypeptide chains, each pair having one "light" (about 25 kDa) and one "heavy" chain (about 50-70 kDa). The N-terminus of each chain defines a variable region of about 100 to 110 or more amino acids primarily responsible for antigen recognition. The terms variable light chain (V_L) and variable heavy chain (V_H) refer to these light and heavy chains respectively.

Antibodies exist, e.g., as intact immunoglobulins or as a number of well-characterized fragments produced by digestion with various peptidases. Thus, for example, pepsin digests an antibody below the disulfide linkages in the hinge region to produce $F(ab)'_2$, a dimer of Fab which itself is a light chain joined to V_H-C_H1 by a disulfide bond. The $F(ab)'_2$ may be reduced under mild conditions to break the disulfide linkage in the hinge region, thereby converting the $F(ab)'_2$ dimer into an Fab' monomer. The Fab' monomer is essentially Fab with part of the hinge region (*see Fundamental Immunology* (Paul ed., 3d ed. 1993). While various antibody fragments are defined in terms of the digestion of an intact antibody, one of skill will appreciate that such fragments may be synthesized *de novo* either chemically or by using recombinant DNA methodology. Thus, the term antibody, as used herein, also includes antibody fragments either produced by the modification of whole antibodies, or those synthesized *de novo* using recombinant DNA methodologies (e.g., single chain Fv) or those identified using phage display libraries (*see, e.g.*, McCafferty *et al., Nature* 348:552-554 (1990)).

For preparation of monoclonal or polyclonal antibodies, any technique known in the art can be used (*see, e.g.*, Kohler & Milstein, *Nature* 256:495-497 (1975); Kozbor *et al., Immunology Today* 4: 72 (1983); Cole *et al.*, pp. 77-96 in *Monoclonal Antibodies and Cancer Therapy* (1985)). Techniques for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms such as other mammals, may be used to express humanized antibodies. Alternatively, phage display technology can be used to identify antibodies and heteromeric Fab fragments that specifically bind to selected antigens (*see, e.g.*, McCafferty *et al., Nature* 348:552-554 (1990); Marks *et al., Biotechnology* 10:779-783 (1992)).

A "chimeric antibody" is an antibody molecule in which (a) the constant region, or a portion thereof, is altered, replaced or exchanged so that the antigen binding

site (variable region) is linked to a constant region of a different or altered class, effector function and/or species, or an entirely different molecule which confers new properties to the chimeric antibody, e.g., an enzyme, toxin, hormone, growth factor, drug, etc.; or (b) the variable region, or a portion thereof, is altered, replaced or exchanged with a variable region having a different or altered antigen specificity.

An "anti-GPCR-B3" antibody is an antibody or antibody fragment that specifically binds a polypeptide encoded by the GPCR-B3 gene, cDNA, or a subsequence thereof.

The term "immunoassay" is an assay that uses an antibody to specifically bind an antigen. The immunoassay is characterized by the use of specific binding properties of a particular antibody to isolate, target, and/or quantify the antigen.

The phrase "specifically (or selectively) binds" to an antibody or "specifically (or selectively) immunoreactive with," when referring to a protein or peptide, refers to a binding reaction that is determinative of the presence of the protein in a heterogeneous population of proteins and other biologics. Thus, under designated immunoassay conditions, the specified antibodies bind to a particular protein at least two times the background and do not substantially bind in a significant amount to other proteins present in the sample. Specific binding to an antibody under such conditions may require an antibody that is selected for its specificity for a particular protein. For example, polyclonal antibodies raised to GPCR-B3 from specific species such as rat, mouse, or human can be selected to obtain only those polyclonal antibodies that are specifically immunoreactive with GPCR-B3 and not with other proteins, except for polymorphic variants and alleles of GPCR-B3. This selection may be achieved by subtracting out antibodies that cross-react with GPCR-B3 molecules from other species. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays are routinely used to select antibodies specifically immunoreactive with a protein (*see, e.g.*, Harlow & Lane, *Antibodies, A Laboratory Manual* (1988), for a description of immunoassay formats and conditions that can be used to determine specific immunoreactivity). Typically a specific or selective reaction will be at least twice background signal or noise and more typically more than 10 to 100 times background.

The phrase "selectively associates with" refers to the ability of a nucleic acid to "selectively hybridize" with another as defined above, or the ability of an antibody to "selectively (or specifically) bind to a protein, as defined above.

By "host cell" is meant a cell that contains an expression vector and supports the replication or expression of the expression vector. Host cells may be prokaryotic cells such as *E. coli*, or eukaryotic cells such as yeast, insect, amphibian, or mammalian cells such as CHO, HeLa and the like, e.g., cultured cells, explants, and cells *in vivo*.

III. Isolation of the nucleic acid encoding GPCR-B3

A. General recombinant DNA methods

This invention relies on routine techniques in the field of recombinant genetics. Basic texts disclosing the general methods of use in this invention include Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (2nd ed. 1989); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); and *Current Protocols in Molecular Biology* (Ausubel *et al.*, eds., 1994)).

For nucleic acids, sizes are given in either kilobases (kb) or base pairs (bp). These are estimates derived from agarose or acrylamide gel electrophoresis, from sequenced nucleic acids, or from published DNA sequences. For proteins, sizes are given in kilodaltons (kDa) or amino acid residue numbers. Proteins sizes are estimated from gel electrophoresis, from sequenced proteins, from derived amino acid sequences, or from published protein sequences.

Oligonucleotides that are not commercially available can be chemically synthesized according to the solid phase phosphoramidite triester method first described by Beaucage & Caruthers, *Tetrahedron Letts*. 22:1859-1862 (1981), using an automated synthesizer, as described in Van Devanter *et. al.*, *Nucleic Acids Res*. 12:6159-6168 (1984). Purification of oligonucleotides is by either native acrylamide gel electrophoresis or by anion-exchange HPLC as described in Pearson & Reanier, *J. Chrom*. 255:137-149 (1983).

The sequence of the cloned genes and synthetic oligonucleotides can be verified after cloning using, e.g., the chain termination method for sequencing double-stranded templates of Wallace *et al., Gene* 16:21-26 (1981).

B. Cloning methods for the isolation of nucleotide sequences encoding GPCR-B3

In general, the nucleic acid sequences encoding GPCR-B3 and related nucleic acid sequence homologs are cloned from cDNA and genomic DNA libraries by

hybridization with a probe, or isolated using amplification techniques with oligonucleotide primers. For example, GPCR-B3 sequences are typically isolated from mammalian nucleic acid (genomic or cDNA) libraries by hybridizing with a nucleic acid probe, the sequence of which can be derived from SEQ ID NOS:4-6. A suitable tissue from which GPCR-B3 RNA and cDNA can be isolated is tongue tissue, optionally taste bud tissue or individual taste cells.

Amplification techniques using primers can also be used to amplify and isolate GPCR-B3 from DNA or RNA. The degenerate primers encoding the following amino acid sequences can also be used to amplify a sequence of GPCR-B3: SEQ ID NOS:7-8 (*see, e.g.*, Dieffenfach & Dveksler, *PCR Primer: A Laboratory Manual* (1995)). These primers can be used, e.g., to amplify either the full length sequence or a probe of one to several hundred nucleotides, which is then used to screen a mammalian library for full-length GPCR-B3.

Nucleic acids encoding GPCR-B3 can also be isolated from expression libraries using antibodies as probes. Such polyclonal or monoclonal antibodies can be raised using the sequence of SEQ ID NOS:1-3.

GPCR-B3 polymorphic variants, alleles, and interspecies homologs that are substantially identical to GPCR-B3 can be isolated using GPCR-B3 nucleic acid probes, and oligonucleotides under stringent hybridization conditions, by screening libraries. Alternatively, expression libraries can be used to clone GPCR-B3 and GPCR-B3 polymorphic variants, alleles, and interspecies homologs, by detecting expressed homologs immunologically with antisera or purified antibodies made against GPCR-B3, which also recognize and selectively bind to the GPCR-B3 homolog.

To make a cDNA library, one should choose a source that is rich in GPCR-B3 mRNA, e.g., tongue tissue, or isolated taste buds. The mRNA is then made into cDNA using reverse transcriptase, ligated into a recombinant vector, and transfected into a recombinant host for propagation, screening and cloning. Methods for making and screening cDNA libraries are well known (*see, e.g.*, Gubler & Hoffman, *Gene* 25:263-269 (1983); Sambrook *et al.*, *supra*; Ausubel *et al.*, *supra*).

For a genomic library, the DNA is extracted from the tissue and either mechanically sheared or enzymatically digested to yield fragments of about 12-20 kb. The fragments are then separated by gradient centrifugation from undesired sizes and are constructed in bacteriophage lambda vectors. These vectors and phage are packaged *in vitro*. Recombinant phage are analyzed by plaque hybridization as described in Benton &

Davis, *Science* 196:180-182 (1977). Colony hybridization is carried out as generally described in Grunstein *et al., Proc. Natl. Acad. Sci. USA.,* 72:3961-3965 (1975).

An alternative method of isolating GPCR-B3 nucleic acid and its homologs combines the use of synthetic oligonucleotide primers and amplification of an RNA or DNA template (*see* U.S. Patents 4,683,195 and 4,683,202; *PCR Protocols: A Guide to Methods and Applications* (Innis *et al.*, eds, 1990)). Methods such as polymerase chain reaction (PCR) and ligase chain reaction (LCR) can be used to amplify nucleic acid sequences of GPCR-B3 directly from mRNA, from cDNA, from genomic libraries or cDNA libraries. Degenerate oligonucleotides can be designed to amplify GPCR-B3 homologs using the sequences provided herein. Restriction endonuclease sites can be incorporated into the primers. Polymerase chain reaction or other *in vitro* amplification methods may also be useful, for example, to clone nucleic acid sequences that code for proteins to be expressed, to make nucleic acids to use as probes for detecting the presence of GPCR-B3 encoding mRNA in physiological samples, for nucleic acid sequencing, or for other purposes. Genes amplified by the PCR reaction can be purified from agarose gels and cloned into an appropriate vector.

Gene expression of GPCR-B3 can also be analyzed by techniques known in the art, e.g., reverse transcription and amplification of mRNA, isolation of total RNA or poly A^+ RNA, northern blotting, dot blotting, *in situ* hybridization, RNase protection, probing DNA microchip arrays, and the like. In one embodiment, high density oligonucleotide analysis technology (e.g., GeneChip™) is used to identify homologs and polymorphic variants of the GPCRs of the invention. In the case where the homologs being identified are linked to a known disease, they can be used with GeneChip™ as a diagnostic tool in detecting the disease in a biological sample, *see, e.g.*, Gunthand *et al., AIDS Res. Hum. Retroviruses* 14: 869-876 (1998); Kozal *et al., Nat. Med.* 2:753-759 (1996); Matson *et al., Anal. Biochem.* 224:110-106 (1995); Lockhart *et al., Nat. Biotechnol.* 14:1675-1680 (1996); Gingeras *et al., Genome Res.* 8:435-448 (1998); Hacia *et al., Nucleic Acids Res.* 26:3865-3866 (1998).

Synthetic oligonucleotides can be used to construct recombinant GPCR-B3 genes for use as probes or for expression of protein. This method is performed using a series of overlapping oligonucleotides usually 40-120 bp in length, representing both the sense and non-sense strands of the gene. These DNA fragments are then annealed, ligated and cloned. Alternatively, amplification techniques can be used with precise

primers to amplify a specific subsequence of the GPCR-B3 nucleic acid. The specific subsequence is then ligated into an expression vector.

The nucleic acid encoding GPCR-B3 is typically cloned into intermediate vectors before transformation into prokaryotic or eukaryotic cells for replication and/or expression. These intermediate vectors are typically prokaryote vectors, e.g., plasmids, or shuttle vectors.

Optionally, nucleic acids encoding chimeric proteins comprising GPCR-B3 or domains thereof can be made according to standard techniques. For example, a domain such as ligand binding domain, an extracellular domain, a transmembrane domain (e.g., one comprising seven transmembrane regions and cytosolic loops), the transmembrane domain and a cytoplasmic domain, an active site, a subunit association region, etc., can be covalently linked to a heterologous protein. For example, an extracellular domain can be linked to a heterologous GPCR transmembrane domain, or a heterologous GPCR extracellular domain can be linked to a transmembrane domain. Other heterologous proteins of choice include, e.g., green fluorescent protein, β-gal, glutamate receptor, and the rhodopsin presequence.

C. Expression in prokaryotes and eukaryotes

To obtain high level expression of a cloned gene or nucleic acid, such as those cDNAs encoding GPCR-B3, one typically subclones GPCR-B3 into an expression vector that contains a strong promoter to direct transcription, a transcription/translation terminator, and if for a nucleic acid encoding a protein, a ribosome binding site for translational initiation. Suitable bacterial promoters are well known in the art and described, e.g., in Sambrook *et al.* and Ausubel *et al.* Bacterial expression systems for expressing the GPCR-B3 protein are available in, e.g., *E. coli, Bacillus sp.,* and *Salmonella* (Palva *et al., Gene* 22:229-235 (1983); Mosbach *et al., Nature* 302:543-545 (1983). Kits for such expression systems are commercially available. Eukaryotic expression systems for mammalian cells, yeast, and insect cells are well known in the art and are also commercially available. In one embodiment, the eukaryotic expression vector is an adenoviral vector, an adeno-associated vector, or a retroviral vector.

The promoter used to direct expression of a heterologous nucleic acid depends on the particular application. The promoter is optionally positioned about the same distance from the heterologous transcription start site as it is from the transcription

start site in its natural setting. As is known in the art, however, some variation in this distance can be accommodated without loss of promoter function.

In addition to the promoter, the expression vector typically contains a transcription unit or expression cassette that contains all the additional elements required for the expression of the GPCR-B3 encoding nucleic acid in host cells. A typical expression cassette thus contains a promoter operably linked to the nucleic acid sequence encoding GPCR-B3 and signals required for efficient polyadenylation of the transcript, ribosome binding sites, and translation termination. The nucleic acid sequence encoding GPCR-B3 may typically be linked to a cleavable signal peptide sequence to promote secretion of the encoded protein by the transformed cell. Such signal peptides would include, among others, the signal peptides from tissue plasminogen activator, insulin, and neuron growth factor, and juvenile hormone esterase of *Heliothis virescens*. Additional elements of the cassette may include enhancers and, if genomic DNA is used as the structural gene, introns with functional splice donor and acceptor sites.

In addition to a promoter sequence, the expression cassette should also contain a transcription termination region downstream of the structural gene to provide for efficient termination. The termination region may be obtained from the same gene as the promoter sequence or may be obtained from different genes.

The particular expression vector used to transport the genetic information into the cell is not particularly critical. Any of the conventional vectors used for expression in eukaryotic or prokaryotic cells may be used. Standard bacterial expression vectors include plasmids such as pBR322 based plasmids, pSKF, pET23D, and fusion expression systems such as GST and LacZ. Epitope tags can also be added to recombinant proteins to provide convenient methods of isolation, e.g., c-myc.

Expression vectors containing regulatory elements from eukaryotic viruses are typically used in eukaryotic expression vectors, e.g., SV40 vectors, papilloma virus vectors, and vectors derived from Epstein-Barr virus. Other exemplary eukaryotic vectors include pMSG, pAV009/A$^+$, pMTO10/A$^+$, pMAMneo-5, baculovirus pDSVE, and any other vector allowing expression of proteins under the direction of the SV40 early promoter, SV40 later promoter, metallothionein promoter, murine mammary tumor virus promoter, Rous sarcoma virus promoter, polyhedrin promoter, or other promoters shown effective for expression in eukaryotic cells.

Some expression systems have markers that provide gene amplification such as thymidine kinase, hygromycin B phosphotransferase, and dihydrofolate reductase.

Alternatively, high yield expression systems not involving gene amplification are also suitable, such as using a baculovirus vector in insect cells, with a GPCR-B3 encoding sequence under the direction of the polyhedrin promoter or other strong baculovirus promoters.

The elements that are typically included in expression vectors also include a replicon that functions in *E. coli*, a gene encoding antibiotic resistance to permit selection of bacteria that harbor recombinant plasmids, and unique restriction sites in nonessential regions of the plasmid to allow insertion of eukaryotic sequences. The particular antibiotic resistance gene chosen is not critical, any of the many resistance genes known in the art are suitable. The prokaryotic sequences are typically chosen such that they do not interfere with the replication of the DNA in eukaryotic cells, if necessary.

Standard transfection methods are used to produce bacterial, mammalian, yeast or insect cell lines that express large quantities of GPCR-B3 protein, which are then purified using standard techniques (*see, e.g.*, Colley *et al.*, *J. Biol. Chem.* 264:17619-17622 (1989); *Guide to Protein Purification*, in *Methods in Enzymology*, vol. 182 (Deutscher, ed., 1990)). Transformation of eukaryotic and prokaryotic cells are performed according to standard techniques (*see, e.g.*, Morrison, *J. Bact.* 132:349-351 (1977); Clark-Curtiss & Curtiss, *Methods in Enzymology* 101:347-362 (Wu *et al.*, eds, 1983).

Any of the well known procedures for introducing foreign nucleotide sequences into host cells may be used. These include the use of calcium phosphate transfection, polybrene, protoplast fusion, electroporation, liposomes, microinjection, plasma vectors, viral vectors and any of the other well known methods for introducing cloned genomic DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell (*see, e.g.*, Sambrook *et al.*, *supra*). It is only necessary that the particular genetic engineering procedure used be capable of successfully introducing at least one gene into the host cell capable of expressing GPCR-B3.

After the expression vector is introduced into the cells, the transfected cells are cultured under conditions favoring expression of GPCR-B3, which is recovered from the culture using standard techniques identified below.

IV. Purification of GPCR-B3

Either naturally occurring or recombinant GPCR-B3 can be purified for use in functional assays. Optionally, recombinant GPCR-B3 is purified. Naturally

occurring GPCR-B3 is purified, e.g., from mammalian tissue such as tongue tissue, and any other source of a GPCR-B3 homolog. Recombinant GPCR-B3 is purified from any suitable expression system, e.g., bacterial and eukaryotic expression systems, e.g., CHO cells or insect cells.

GPCR-B3 may be purified to substantial purity by standard techniques, including selective precipitation with such substances as ammonium sulfate; column chromatography, immunopurification methods, and others (*see, e.g.*, Scopes, *Protein Purification: Principles and Practice* (1982); U.S. Patent No. 4,673,641; Ausubel *et al., supra*; and Sambrook *et al., supra*).

A number of procedures can be employed when recombinant GPCR-B3 is being purified. For example, proteins having established molecular adhesion properties can be reversible fused to GPCR-B3. With the appropriate ligand, GPCR-B3 can be selectively adsorbed to a purification column and then freed from the column in a relatively pure form. The fused protein is then removed by enzymatic activity. Finally GPCR-B3 could be purified using immunoaffinity columns.

A. Purification of GPCR-B3 from recombinant cells

Recombinant proteins are expressed by transformed bacteria or eukaryotic cells such as CHO or insect cells in large amounts, typically after promoter induction; but expression can be constitutive. Promoter induction with IPTG is a one example of an inducible promoter system. Cells are grown according to standard procedures in the art. Fresh or frozen cells are used for isolation of protein.

Proteins expressed in bacteria may form insoluble aggregates ("inclusion bodies"). Several protocols are suitable for purification of GPCR-B3 inclusion bodies. For example, purification of inclusion bodies typically involves the extraction, separation and/or purification of inclusion bodies by disruption of bacterial cells, e.g., by incubation in a buffer of 50 mM TRIS/HCL pH 7.5, 50 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 0.1 mM ATP, and 1 mM PMSF. The cell suspension can be lysed using 2-3 passages through a French Press, homogenized using a Polytron (Brinkman Instruments) or sonicated on ice. Alternate methods of lysing bacteria are apparent to those of skill in the art (*see, e.g.*, Sambrook *et al., supra*; Ausubel *et al., supra*).

If necessary, the inclusion bodies are solubilized, and the lysed cell suspension is typically centrifuged to remove unwanted insoluble matter. Proteins that formed the inclusion bodies may be renatured by dilution or dialysis with a compatible

buffer. Suitable solvents include, but are not limited to urea (from about 4 M to about 8 M), formamide (at least about 80%, volume/volume basis), and guanidine hydrochloride (from about 4 M to about 8 M). Some solvents which are capable of solubilizing aggregate-forming proteins, for example SDS (sodium dodecyl sulfate), 70% formic acid, are inappropriate for use in this procedure due to the possibility of irreversible denaturation of the proteins, accompanied by a lack of immunogenicity and/or activity. Although guanidine hydrochloride and similar agents are denaturants, this denaturation is not irreversible and renaturation may occur upon removal (by dialysis, for example) or dilution of the denaturant, allowing re-formation of immunologically and/or biologically active protein. Other suitable buffers are known to those skilled in the art. GPCR-B3 is separated from other bacterial proteins by standard separation techniques, e.g., with Ni-NTA agarose resin.

Alternatively, it is possible to purify GPCR-B3 from bacteria periplasm. After lysis of the bacteria, when GPCR-B3 is exported into the periplasm of the bacteria, the periplasmic fraction of the bacteria can be isolated by cold osmotic shock in addition to other methods known to skill in the art. To isolate recombinant proteins from the periplasm, the bacterial cells are centrifuged to form a pellet. The pellet is resuspended in a buffer containing 20% sucrose. To lyse the cells, the bacteria are centrifuged and the pellet is resuspended in ice-cold 5 mM $MgSO_4$ and kept in an ice bath for approximately 10 minutes. The cell suspension is centrifuged and the supernatant decanted and saved. The recombinant proteins present in the supernatant can be separated from the host proteins by standard separation techniques well known to those of skill in the art.

B. Standard protein separation techniques for purifying GPCR-B3

Solubility fractionation

Often as an initial step, particularly if the protein mixture is complex, an initial salt fractionation can separate many of the unwanted host cell proteins (or proteins derived from the cell culture media) from the recombinant protein of interest. In one embodiment, the salt is ammonium sulfate. Ammonium sulfate precipitates proteins by effectively reducing the amount of water in the protein mixture. Proteins then precipitate on the basis of their solubility. The more hydrophobic a protein is, the more likely it is to precipitate at lower ammonium sulfate concentrations. A typical protocol includes adding saturated ammonium sulfate to a protein solution so that the resultant ammonium sulfate concentration is between 20-30%. This concentration will precipitate the most

hydrophobic of proteins. The precipitate is then discarded (unless the protein of interest is hydrophobic) and ammonium sulfate is added to the supernatant to a concentration known to precipitate the protein of interest. The precipitate is then solubilized in buffer and the excess salt removed if necessary, either through dialysis or diafiltration. Other methods that rely on solubility of proteins, such as cold ethanol precipitation, are well known to those of skill in the art and can be used to fractionate complex protein mixtures.

Size differential filtration

The molecular weight of GPCR-B3 can be used to isolated it from proteins of greater and lesser size using ultrafiltration through membranes of different pore size (for example, Amicon or Millipore membranes). As a first step, the protein mixture is ultrafiltered through a membrane with a pore size that has a lower molecular weight cut-off than the molecular weight of the protein of interest. The retentate of the ultrafiltration is then ultrafiltered against a membrane with a molecular cut off greater than the molecular weight of the protein of interest. The recombinant protein will pass through the membrane into the filtrate. The filtrate can then be chromatographed as described below.

Column chromatography

GPCR-B3 can also be separated from other proteins on the basis of its size, net surface charge, hydrophobicity, and affinity for ligands. In addition, antibodies raised against proteins can be conjugated to column matrices and the proteins immunopurified. All of these methods are well known in the art. It will be apparent to one of skill that chromatographic techniques can be performed at any scale and using equipment from many different manufacturers (e.g., Pharmacia Biotech).

V. Immunological detection of GPCR-B3

In addition to the detection of GPCR-B3 genes and gene expression using nucleic acid hybridization technology, one can also use immunoassays to detect GPCR-B3, e.g., to identify taste receptor cells and variants of GPCR-B3. Immunoassays can be used to qualitatively or quantitatively analyze GPCR-B3. A general overview of the applicable technology can be found in Harlow & Lane, *Antibodies: A Laboratory Manual* (1988).

A. Antibodies to GPCR-B3

Methods of producing polyclonal and monoclonal antibodies that react specifically with GPCR-B3 are known to those of skill in the art (*see, e.g.*, Coligan, *Current Protocols in Immunology* (1991); Harlow & Lane, *supra*; Goding, *Monoclonal Antibodies: Principles and Practice* (2d ed. 1986); and Kohler & Milstein, *Nature* 256:495-497 (1975). Such techniques include antibody preparation by selection of antibodies from libraries of recombinant antibodies in phage or similar vectors, as well as preparation of polyclonal and monoclonal antibodies by immunizing rabbits or mice (*see, e.g.*, Huse *et al.*, *Science* 246:1275-1281 (1989); Ward *et al.*, *Nature* 341:544-546 (1989)).

A number of GPCR-B3 comprising immunogens may be used to produce antibodies specifically reactive with GPCR-B3. For example, recombinant GPCR-B3 or an antigenic fragment thereof, is isolated as described herein. Recombinant protein can be expressed in eukaryotic or prokaryotic cells as described above, and purified as generally described above. Recombinant protein is one embodiment of an immunogen for the production of monoclonal or polyclonal antibodies. Alternatively, a synthetic peptide derived from the sequences disclosed herein and conjugated to a carrier protein can be used an immunogen. Naturally occurring protein may also be used either in pure or impure form. The product is then injected into an animal capable of producing antibodies. Either monoclonal or polyclonal antibodies may be generated, for subsequent use in immunoassays to measure the protein.

Methods of production of polyclonal antibodies are known to those of skill in the art. An inbred strain of mice (e.g., BALB/C mice) or rabbits is immunized with the protein using a standard adjuvant, such as Freund's adjuvant, and a standard immunization protocol. The animal's immune response to the immunogen preparation is monitored by taking test bleeds and determining the titer of reactivity to GPCR-B3. When appropriately high titers of antibody to the immunogen are obtained, blood is collected from the animal and antisera are prepared. Further fractionation of the antisera to enrich for antibodies reactive to the protein can be done if desired (*see* Harlow & Lane, *supra*).

Monoclonal antibodies may be obtained by various techniques familiar to those skilled in the art. Briefly, spleen cells from an animal immunized with a desired antigen are immortalized, commonly by fusion with a myeloma cell (*see* Kohler & Milstein, *Eur. J. Immunol.* 6:511-519 (1976)). Alternative methods of immortalization

include transformation with Epstein Barr Virus, oncogenes, or retroviruses, or other methods well known in the art. Colonies arising from single immortalized cells are screened for production of antibodies of the desired specificity and affinity for the antigen, and yield of the monoclonal antibodies produced by such cells may be enhanced by various techniques, including injection into the peritoneal cavity of a vertebrate host. Alternatively, one may isolate DNA sequences which encode a monoclonal antibody or a binding fragment thereof by screening a DNA library from human B cells according to the general protocol outlined by Huse *et al.*, *Science* 246:1275-1281 (1989).

Monoclonal antibodies and polyclonal sera are collected and titered against the immunogen protein in an immunoassay, for example, a solid phase immunoassay with the immunogen immobilized on a solid support. Typically, polyclonal antisera with a titer of 10^4 or greater are selected and tested for their cross reactivity against non-GPCR-B3 proteins or even other related proteins from other organisms, using a competitive binding immunoassay. Specific polyclonal antisera and monoclonal antibodies will usually bind with a K_d of at least about 0.1 mM, more usually at least about 1 μM, optionally at least about 0.1 μM or better, and optionally 0.01 μM or better.

Once GPCR-B3 specific antibodies are available, GPCR-B3 can be detected by a variety of immunoassay methods. For a review of immunological and immunoassay procedures, see *Basic and Clinical Immunology* (Stites & Terr eds., 7th ed. 1991). Moreover, the immunoassays of the present invention can be performed in any of several configurations, which are reviewed extensively in *Enzyme Immunoassay* (Maggio, ed., 1980); and Harlow & Lane, *supra*.

B. Immunological binding assays

GPCR-B3 can be detected and/or quantified using any of a number of well recognized immunological binding assays (*see, e.g.*, U.S. Patents 4,366,241; 4,376,110; 4,517,288; and 4,837,168). For a review of the general immunoassays, see also *Methods in Cell Biology: Antibodies in Cell Biology*, volume 37 (Asai, ed. 1993); *Basic and Clinical Immunology* (Stites & Terr, eds., 7th ed. 1991). Immunological binding assays (or immunoassays) typically use an antibody that specifically binds to a protein or antigen of choice (in this case the GPCR-B3 or antigenic subsequence thereof). The antibody (e.g., anti-GPCR-B3) may be produced by any of a number of means well known to those of skill in the art and as described above.

Immunoassays also often use a labeling agent to specifically bind to and label the complex formed by the antibody and antigen. The labeling agent may itself be one of the moieties comprising the antibody/antigen complex. Thus, the labeling agent may be a labeled GPCR-B3 polypeptide or a labeled anti-GPCR-B3 antibody. Alternatively, the labeling agent may be a third moiety, such a secondary antibody, that specifically binds to the antibody/GPCR-B3 complex (a secondary antibody is typically specific to antibodies of the species from which the first antibody is derived). Other proteins capable of specifically binding immunoglobulin constant regions, such as protein A or protein G may also be used as the label agent. These proteins exhibit a strong non-immunogenic reactivity with immunoglobulin constant regions from a variety of species (*see, e.g.*, Kronval *et al.*, *J. Immunol.* 111:1401-1406 (1973); Akerstrom *et al.*, *J. Immunol.* 135:2589-2542 (1985)). The labeling agent can be modified with a detectable moiety, such as biotin, to which another molecule can specifically bind, such as streptavidin. A variety of detectable moieties are well known to those skilled in the art.

Throughout the assays, incubation and/or washing steps may be required after each combination of reagents. Incubation steps can vary from about 5 seconds to several hours, optionally from about 5 minutes to about 24 hours. However, the incubation time will depend upon the assay format, antigen, volume of solution, concentrations, and the like. Usually, the assays will be carried out at ambient temperature, although they can be conducted over a range of temperatures, such as 10°C to 40°C.

<u>Non-competitive assay formats</u>

Immunoassays for detecting GPCR-B3 in samples may be either competitive or noncompetitive. Noncompetitive immunoassays are assays in which the amount of antigen is directly measured. In one embodiment "sandwich" assay, for example, the anti-GPCR-B3 antibodies can be bound directly to a solid substrate on which they are immobilized. These immobilized antibodies then capture GPCR-B3 present in the test sample. GPCR-B3 is thus immobilized is then bound by a labeling agent, such as a second GPCR-B3 antibody bearing a label. Alternatively, the second antibody may lack a label, but it may, in turn, be bound by a labeled third antibody specific to antibodies of the species from which the second antibody is derived. The second or third antibody is typically modified with a detectable moiety, such as biotin, to

which another molecule specifically binds, e.g., streptavidin, to provide a detectable moiety.

Competitive assay formats

In competitive assays, the amount of GPCR-B3 present in the sample is measured indirectly by measuring the amount of a known, added (exogenous) GPCR-B3 displaced (competed away) from an anti-GPCR-B3 antibody by the unknown GPCR-B3 present in a sample. In one competitive assay, a known amount of GPCR-B3 is added to a sample and the sample is then contacted with an antibody that specifically binds to GPCR-B3. The amount of exogenous GPCR-B3 bound to the antibody is inversely proportional to the concentration of GPCR-B3 present in the sample. In one embodiment, the antibody is immobilized on a solid substrate. The amount of GPCR-B3 bound to the antibody may be determined either by measuring the amount of GPCR-B3 present in a GPCR-B3/antibody complex, or alternatively by measuring the amount of remaining uncomplexed protein. The amount of GPCR-B3 may be detected by providing a labeled GPCR-B3 molecule.

A hapten inhibition assay is another competitive assay. In this assay the known GPCR-B3, is immobilized on a solid substrate. A known amount of anti-GPCR-B3 antibody is added to the sample, and the sample is then contacted with the immobilized GPCR-B3. The amount of anti-GPCR-B3 antibody bound to the known immobilized GPCR-B3 is inversely proportional to the amount of GPCR-B3 present in the sample. Again, the amount of immobilized antibody may be detected by detecting either the immobilized fraction of antibody or the fraction of the antibody that remains in solution. Detection may be direct where the antibody is labeled or indirect by the subsequent addition of a labeled moiety that specifically binds to the antibody as described above.

Cross-reactivity determinations

Immunoassays in the competitive binding format can also be used for crossreactivity determinations. For example, a protein at least partially encoded by SEQ ID NOS:1-3 can be immobilized to a solid support. Proteins (e.g., GPCR-B3 proteins and homologs) are added to the assay that compete for binding of the antisera to the immobilized antigen. The ability of the added proteins to compete for binding of the antisera to the immobilized protein is compared to the ability of GPCR-B3 encoded by

SEQ ID NO:1-3 to compete with itself. The percent crossreactivity for the above proteins is calculated, using standard calculations. Those antisera with less than 10% crossreactivity with each of the added proteins listed above are selected and pooled. The cross-reacting antibodies are optionally removed from the pooled antisera by immunoabsorption with the added considered proteins, e.g., distantly related homologs.

The immunoabsorbed and pooled antisera are then used in a competitive binding immunoassay as described above to compare a second protein, thought to be perhaps an allele or polymorphic variant of GPCR-B3, to the immunogen protein (i.e., GPCR-B3 of SEQ ID NOS:1-3). In order to make this comparison, the two proteins are each assayed at a wide range of concentrations and the amount of each protein required to inhibit 50% of the binding of the antisera to the immobilized protein is determined. If the amount of the second protein required to inhibit 50% of binding is less than 10 times the amount of the protein encoded by SEQ ID NOS:1-3 that is required to inhibit 50% of binding, then the second protein is said to specifically bind to the polyclonal antibodies generated to a GPCR-B3 immunogen.

Other assay formats

Western blot (immunoblot) analysis is used to detect and quantify the presence of GPCR-B3 in the sample. The technique generally comprises separating sample proteins by gel electrophoresis on the basis of molecular weight, transferring the separated proteins to a suitable solid support, (such as a nitrocellulose filter, a nylon filter, or derivatized nylon filter), and incubating the sample with the antibodies that specifically bind GPCR-B3. The anti-GPCR-B3 antibodies specifically bind to the GPCR-B3 on the solid support. These antibodies may be directly labeled or alternatively may be subsequently detected using labeled antibodies (e.g., labeled sheep anti-mouse antibodies) that specifically bind to the anti-GPCR-B3 antibodies.

Other assay formats include liposome immunoassays (LIA), which use liposomes designed to bind specific molecules (e.g., antibodies) and release encapsulated reagents or markers. The released chemicals are then detected according to standard techniques (*see* Monroe *et al.*, *Amer. Clin. Prod. Rev.* 5:34-41 (1986)).

Reduction of non-specific binding

One of skill in the art will appreciate that it is often desirable to minimize non-specific binding in immunoassays. Particularly, where the assay involves an antigen

or antibody immobilized on a solid substrate it is desirable to minimize the amount of non-specific binding to the substrate. Means of reducing such non-specific binding are well known to those of skill in the art. Typically, this technique involves coating the substrate with a proteinaceous composition. In particular, protein compositions such as bovine serum albumin (BSA), nonfat powdered milk, and gelatin are widely used.

Labels

The particular label or detectable group used in the assay is not a critical aspect of the invention, as long as it does not significantly interfere with the specific binding of the antibody used in the assay. The detectable group can be any material having a detectable physical or chemical property. Such detectable labels have been well-developed in the field of immunoassays and, in general, most any label useful in such methods can be applied to the present invention. Thus, a label is any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Useful labels in the present invention include magnetic beads (e.g., DYNABEADS™), fluorescent dyes (e.g., fluorescein isothiocyanate, Texas red, rhodamine, and the like), radiolabels (e.g., ^{3}H, ^{125}I, ^{35}S, ^{14}C, or ^{32}P), enzymes (e.g., horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic beads (e.g., polystyrene, polypropylene, latex, etc.).

The label may be coupled directly or indirectly to the desired component of the assay according to methods well known in the art. As indicated above, a wide variety of labels may be used, with the choice of label depending on sensitivity required, ease of conjugation with the compound, stability requirements, available instrumentation, and disposal provisions.

Non-radioactive labels are often attached by indirect means. Generally, a ligand molecule (e.g., biotin) is covalently bound to the molecule. The ligand then binds to another molecules (e.g., streptavidin) molecule, which is either inherently detectable or covalently bound to a signal system, such as a detectable enzyme, a fluorescent compound, or a chemiluminescent compound. The ligands and their targets can be used in any suitable combination with antibodies that recognize GPCR-B3, or secondary antibodies that recognize anti-GPCR-B3.

The molecules can also be conjugated directly to signal generating compounds, e.g., by conjugation with an enzyme or fluorophore. Enzymes of interest as

labels will primarily be hydrolases, particularly phosphatases, esterases and glycosidases, or oxidotases, particularly peroxidases. Fluorescent compounds include fluorescein and its derivatives, rhodamine and its derivatives, dansyl, umbelliferone, etc. Chemiluminescent compounds include luciferin, and 2,3-dihydrophthalazinediones, e.g., luminol. For a review of various labeling or signal producing systems that may be used, see U.S. Patent No. 4,391,904.

Means of detecting labels are well known to those of skill in the art. Thus, for example, where the label is a radioactive label, means for detection include a scintillation counter or photographic film as in autoradiography. Where the label is a fluorescent label, it may be detected by exciting the fluorochrome with the appropriate wavelength of light and detecting the resulting fluorescence. The fluorescence may be detected visually, by means of photographic film, by the use of electronic detectors such as charge coupled devices (CCDs) or photomultipliers and the like. Similarly, enzymatic labels may be detected by providing the appropriate substrates for the enzyme and detecting the resulting reaction product. Finally simple colorimetric labels may be detected simply by observing the color associated with the label. Thus, in various dipstick assays, conjugated gold often appears pink, while various conjugated beads appear the color of the bead.

Some assay formats do not require the use of labeled components. For instance, agglutination assays can be used to detect the presence of the target antibodies. In this case, antigen-coated particles are agglutinated by samples comprising the target antibodies. In this format, none of the components need be labeled and the presence of the target antibody is detected by simple visual inspection.

VI. Assays for modulators of GPCR-B3

A. Assays for GPCR-B3 activity

GPCR-B3 and its alleles and polymorphic variants are G-protein coupled receptors that participate in taste transduction. The activity of GPCR-B3 polypeptides can be assessed using a variety of *in vitro* and *in vivo* assays that determine functional, physical and chemical effects, e.g., measuring ligand binding (e.g., by radioactive ligand binding), second messengers (e.g., cAMP, cGMP, IP_3, DAG, or Ca^{2+}), ion flux, phosphorylation levels, transcription levels, neurotransmitter levels, and the like. Furthermore, such assays can be used to test for inhibitors and activators of GPCR-B3.

Modulators can also be genetically altered versions of GPCR-B3. Such modulators of taste transduction activity are useful for customizing taste.

The GPCR-B3 of the assay will be selected from a polypeptide having a sequence of SEQ ID NOS:1-3 or conservatively modified variant thereof. Alternatively, the GPCR-B3 of the assay will be derived from a eukaryote and include an amino acid subsequence having amino acid sequence identity SEQ ID NOS:1-3. Generally, the amino acid sequence identity will be at least 70%, optionally at least 85%, most optionally at least 90-95%. Optionally, the polypeptide of the assays will comprise a domain of GPCR-B3, such as an extracellular domain, transmembrane domain, cytoplasmic domain, ligand binding domain, subunit association domain, active site, and the like. Either GPCR-B3 or a domain thereof can be covalently linked to a heterologous protein to create a chimeric protein used in the assays described herein.

Modulators of GPCR-B3 activity are tested using GPCR-B3 polypeptides as described above, either recombinant or naturally occurring. The protein can be isolated, expressed in a cell, expressed in a membrane derived from a cell, expressed in tissue or in an animal, either recombinant or naturally occurring. For example, tongue slices, dissociated cells from a tongue, transformed cells, or membranes can b used.. Modulation is tested using one of the *in vitro* or *in vivo* assays described herein. Taste transduction can also be examined *in vitro* with soluble or solid state reactions, using a chimeric molecule such as an extracellular domain of a receptor covalently linked to a heterologous signal transduction domain, or a heterologous extracellular domain covalently linked to the transmembrane and or cytoplasmic domain of a receptor. Furthermore, ligand-binding domains of the protein of interest can be used in vitro in soluble or solid state reactions to assay for ligand binding.

Ligand binding to GPCR-B3, a domain, or chimeric protein can be tested in solution, in a bilayer membrane, attached to a solid phase, in a lipid monolayer, or in vesicles. Binding of a modulator can be tested using, e.g., changes in spectroscopic characteristics (e.g., fluorescence, absorbance, refractive index) hydrodynamic (e.g., shape), chromatographic, or solubility properties.

Receptor-G-protein interactions can also be examined. For example, binding of the G-protein to the receptor or its release from the receptor can be examined. For example, in the absence of GTP, an activator will lead to the formation of a tight complex of a G protein (all three subunits) with the receptor. This complex can be detected in a variety of ways, as noted above. Such an assay can be modified to search

for inhibitors. Add an activator to the receptor and G protein in the absence of GTP, form a tight complex, and then screen for inhibitors by looking at dissociation of the receptor-G protein complex. In the presence of GTP, release of the alpha subunit of the G protein from the other two G protein subunits serves as a criterion of activation.

An activated or inhibited G-protein will in turn alter the properties of target enzymes, channels, and other effector proteins. The classic examples are the activation of cGMP phosphodiesterase by transducin in the visual system, adenylate cyclase by the stimulatory G-protein, phospholipase C by Gq and other cognate G proteins, and modulation of diverse channels by Gi and other G proteins. Downstream consequences can also be examined such as generation of diacyl glycerol and IP3 by phospholipase C, and in turn, for calcium mobilization by IP3.

Activated GPCR receptors become substrates for kinases that phosphorylate the C-terminal tail of the receptor (and possibly other sites as well). Thus, activators will promote the transfer of ^{32}P from gamma-labeled GTP to the receptor, which can be assayed with a scintillation counter. The phosphorylation of the C-terminal tail will promote the binding of arrestin-like proteins and will interfere with the binding of G-proteins. The kinase/arrestin pathway plays a key role in the desensitization of many GPCR receptors. For example, compounds that modulate the duration a taste receptor stays active would be useful as a means of prolonging a desired taste or cutting off an unpleasant one. For a general review of GPCR signal transduction and methods of assaying signal transduction, *see, e.g., Methods in Enzymology*, vols. 237 and 238 (1994) and volume 96 (1983); Bourne *et al., Nature* 10:349:117-27 (1991); Bourne et al., Nature 348:125-32 (1990); Pitcher *et al., Annu. Rev. Biochem.* 67:653-92 (1998).

Samples or assays that are treated with a potential GPCR-B3 inhibitor or activator are compared to control samples without the test compound, to examine the extent of modulation. Control samples (untreated with activators or inhibitors) are assigned a relative GPCR-B3 activity value of 100. Inhibition of GPCR-B3 is achieved when the GPCR-B3 activity value relative to the control is about 90%, optionally 50%, optionally 25-0%. Activation of GPCR-B3 is achieved when the GPCR-B3 activity value relative to the control is 110%, optionally 150%, 200-500%, or 1000-2000%.

Changes in ion flux may be assessed by determining changes in polarization (i.e., electrical potential) of the cell or membrane expressing GPCR-B3. One means to determine changes in cellular polarization is by measuring changes in current (thereby measuring changes in polarization) with voltage-clamp and patch-clamp

techniques, e.g., the "cell-attached" mode, the "inside-out" mode, and the "whole cell" mode (*see, e.g.*, Ackerman *et al., New Engl. J. Med.* 336:1575-1595 (1997)). Whole cell currents are conveniently determined using the standard methodology (*see, e.g.*, Hamil *et al., PFlugers. Archiv.* 391:85 (1981). Other known assays include: radiolabeled ion flux assays and fluorescence assays using voltage-sensitive dyes (*see, e.g.*, Vestergarrd-Bogind *et al., J. Membrane Biol.* 88:67-75 (1988); Gonzales & Tsien, *Chem. Biol.* 4:269-277 (1997); Daniel *et al., J. Pharmacol. Meth.* 25:185-193 (1991); Holevinsky *et al., J. Membrane Biology* 137:59-70 (1994)). Generally, the compounds to be tested are present in the range from 1 pM to 100 mM.

The effects of the test compounds upon the function of the polypeptides can be measured by examining any of the parameters described above. Any suitable physiological change that affects GPCR activity can be used to assess the influence of a test compound on the polypeptides of this invention. When the functional consequences are determined using intact cells or animals, one can also measure a variety of effects such as transmitter release, hormone release, transcriptional changes to both known and uncharacterized genetic markers (e.g., northern blots), changes in cell metabolism such as cell growth or pH changes, and changes in intracellular second messengers such as Ca^{2+}, IP3 or cAMP.

Assays for G-protein coupled receptors include cells that are loaded with ion or voltage sensitive dyes to report receptor activity. Assays for determining activity of such receptors can also use known agonists and antagonists for other G-protein coupled receptors as negative or positive controls to assess activity of tested compounds. In assays for identifying modulatory compounds (e.g., agonists, antagonists), changes in the level of ions in the cytoplasm or membrane voltage will be monitored using an ion - sensitive or membrane voltage fluorescent indicator, respectively. Among the ion-sensitive indicators and voltage probes that may be employed are those disclosed in the Molecular Probes 1997 Catalog. For G-protein coupled receptors, promiscuous G-proteins such as $G\alpha15$ and $G\alpha16$ can be used in the assay of choice (Wilkie *et al., Proc. Nat'l Acad. Sci. USA* 88:10049-10053 (1991)). Such promiscuous G-proteins allow coupling of a wide range of receptors.

Receptor activation typically initiates subsequent intracellular events, e.g., increases in second messengers such as IP3, which releases intracellular stores of calcium ions. Activation of some G-protein coupled receptors stimulates the formation of inositol

triphosphate (IP3) through phospholipase C-mediated hydrolysis of phosphatidylinositol (Berridge & Irvine, *Nature* 312:315-21 (1984)). IP3 in turn stimulates the release of intracellular calcium ion stores. Thus, a change in cytoplasmic calcium ion levels, or a change in second messenger levels such as IP3 can be used to assess G-protein coupled receptor function. Cells expressing such G-protein coupled receptors may exhibit increased cytoplasmic calcium levels as a result of contribution from both intracellular stores and via activation of ion channels, in which case it may be desirable although not necessary to conduct such assays in calcium-free buffer, optionally supplemented with a chelating agent such as EGTA, to distinguish fluorescence response resulting from calcium release from internal stores.

Other assays can involve determining the activity of receptors which, when activated, result in a change in the level of intracellular cyclic nucleotides, e.g., cAMP or cGMP, by activating or inhibiting enzymes such as adenylate cyclase. There are cyclic nucleotide-gated ion channels, e.g., rod photoreceptor cell channels and olfactory neuron channels that are permeable to cations upon activation by binding of cAMP or cGMP (*see, e.g.*, Altenhofen *et al., Proc. Natl. Acad. Sci. U.S.A.* 88:9868-9872 (1991) and Dhallan *et al., Nature* 347:184-187 (1990)). In cases where activation of the receptor results in a decrease in cyclic nucleotide levels, it may be preferable to expose the cells to agents that increase intracellular cyclic nucleotide levels, e.g., forskolin, prior to adding a receptor-activating compound to the cells in the assay. Cells for this type of assay can be made by co-transfection of a host cell with DNA encoding a cyclic nucleotide-crated ion channel, GPCR phosphatase and DNA encoding a receptor (e.g., certain glutamate receptors, muscarinic acetylcholine receptors, dopamine receptors, serotonin receptors, and the like), which, when activated, causes a change in cyclic nucleotide levels in the cytoplasm.

In one embodiment, GPCR-B3 activity is measured by expressing GPCR-B3 in a heterologous cell with a promiscuous G-protein that links the receptor to a phospholipase C signal transduction pathway (*see* Offermanns & Simon, *J. Biol. Chem.* 270:15175-15180 (1995)). Optionally the cell line is HEK-293 (which does not naturally express GPCR-B3) and the promiscuous G-protein is Gα15 (Offermanns & Simon, *supra*). Modulation of taste transduction is assayed by measuring changes in intracellular Ca^{2+} levels, which change in response to modulation of the GPCR-B3 signal transduction pathway via administration of a molecule that associates with GPCR-B3. Changes in

Ca^{2+} levels are optionally measured using fluorescent Ca^{2+} indicator dyes and fluorometric imaging.

In one embodiment, the changes in intracellular cAMP or cGMP can be measured using immunoassays. The method described in Offermanns & Simon, *J. Biol. Chem.* 270:15175-15180 (1995) may be used to determine the level of cAMP. Also, the method described in Felley-Bosco *et al., Am. J. Resp. Cell and Mol. Biol.* 11:159-164 (1994) may be used to determine the level of cGMP. Further, an assay kit for measuring cAMP and/or cGMP is described in U.S. Patent 4,115,538, herein incorporated by reference.

In another embodiment, phosphatidyl inositol (PI) hydrolysis can be analyzed according to U.S. Patent 5,436,128, herein incorporated by reference. Briefly, the assay involves labeling of cells with ^{3}H-myoinositol for 48 or more hrs. The labeled cells are treated with a test compound for one hour. The treated cells are lysed and extracted in chloroform-methanol-water after which the inositol phosphates were separated by ion exchange chromatography and quantified by scintillation counting. Fold stimulation is determined by calculating the ratio of cpm in the presence of agonist to cpm in the presence of buffer control. Likewise, fold inhibition is determined by calculating the ratio of cpm in the presence of antagonist to cpm in the presence of buffer control (which may or may not contain an agonist).

In another embodiment, transcription levels can be measured to assess the effects of a test compound on signal transduction. A host cell containing the protein of interest is contacted with a test compound for a sufficient time to effect any interactions, and then the level of gene expression is measured. The amount of time to effect such interactions may be empirically determined, such as by running a time course and measuring the level of transcription as a function of time. The amount of transcription may be measured by using any method known to those of skill in the art to be suitable. For example, mRNA expression of the protein of interest may be detected using northern blots or their polypeptide products may be identified using immunoassays. Alternatively, transcription based assays using reporter gene may be used as described in U.S. Patent 5,436,128, herein incorporated by reference. The reporter genes can be, e.g., chloramphenicol acetyltransferase, firefly luciferase, bacterial luciferase, β-galactosidase and alkaline phosphatase. Furthermore, the protein of interest can be used as an indirect

reporter via attachment to a second reporter such as green fluorescent protein (*see, e.g.*, Mistili & Spector, *Nature Biotechnology* 15:961-964 (1997)).

The amount of transcription is then compared to the amount of transcription in either the same cell in the absence of the test compound, or it may be compared with the amount of transcription in a substantially identical cell that lacks the protein of interest. A substantially identical cell may be derived from the same cells from which the recombinant cell was prepared but which had not been modified by introduction of heterologous DNA. Any difference in the amount of transcription indicates that the test compound has in some manner altered the activity of the protein of interest.

B. Modulators

The compounds tested as modulators of GPCR-B3 can be any small chemical compound, or a biological entity, such as a protein, sugar, nucleic acid or lipid. Alternatively, modulators can be genetically altered versions of GPCR-B3. Typically, test compounds will be small chemical molecules and peptides. Essentially any chemical compound can be used as a potential modulator or ligand in the assays of the invention, although most often compounds can be dissolved in aqueous or organic (especially DMSO-based) solutions are used. The assays are designed to screen large chemical libraries by automating the assay steps and providing compounds from any convenient source to assays, which are typically run in parallel (e.g., in microtiter formats on microtiter plates in robotic assays). It will be appreciated that there are many suppliers of chemical compounds, including Sigma (St. Louis, MO), Aldrich (St. Louis, MO), Sigma-Aldrich (St. Louis, MO), Fluka Chemika-Biochemica Analytika (Buchs Switzerland) and the like.

In one embodiment, high throughput screening methods involve providing a combinatorial chemical or peptide library containing a large number of potential therapeutic compounds (potential modulator or ligand compounds). Such "combinatorial chemical libraries" or "ligand libraries" are then screened in one or more assays, as described herein, to identify those library members (particular chemical species or subclasses) that display a desired characteristic activity. The compounds thus identified can serve as conventional "lead compounds" or can themselves be used as potential or actual therapeutics.

A combinatorial chemical library is a collection of diverse chemical compounds generated by either chemical synthesis or biological synthesis, by combining a number of chemical "building blocks" such as reagents. For example, a linear combinatorial chemical library such as a polypeptide library is formed by combining a set of chemical building blocks (amino acids) in every possible way for a given compound length (i.e., the number of amino acids in a polypeptide compound). Millions of chemical compounds can be synthesized through such combinatorial mixing of chemical building blocks.

Preparation and screening of combinatorial chemical libraries is well known to those of skill in the art. Such combinatorial chemical libraries include, but are not limited to, peptide libraries (*see, e.g.,* U.S. Patent 5,010,175, Furka, *Int. J. Pept. Prot. Res.* 37:487-493 (1991) and Houghton *et al., Nature* 354:84-88 (1991)). Other chemistries for generating chemical diversity libraries can also be used. Such chemistries include, but are not limited to: peptoids (e.g., PCT Publication No. WO 91/19735), encoded peptides (e.g., PCT Publication WO 93/20242), random bio-oligomers (e.g., PCT Publication No. WO 92/00091), benzodiazepines (e.g., U.S. Pat. No. 5,288,514), diversomers such as hydantoins, benzodiazepines and dipeptides (Hobbs *et al., Proc. Nat. Acad. Sci. USA* 90:6909-6913 (1993)), vinylogous polypeptides (Hagihara *et al., J. Amer. Chem. Soc.* 114:6568 (1992)), nonpeptidal peptidomimetics with glucose scaffolding (Hirschmann *et al., J. Amer. Chem. Soc.* 114:9217-9218 (1992)), analogous organic syntheses of small compound libraries (Chen *et al., J. Amer. Chem. Soc.* 116:2661 (1994)), oligocarbamates (Cho *et al., Science* 261:1303 (1993)), and/or peptidyl phosphonates (Campbell *et al., J. Org. Chem.* 59:658 (1994)), nucleic acid libraries (*see* Ausubel, Berger and Sambrook, all *supra*), peptide nucleic acid libraries (*see, e.g.,* U.S. Patent 5,539,083), antibody libraries (*see, e.g.,* Vaughn *et al., Nature Biotechnology,* 14(3):309-314 (1996) and PCT/US96/10287), carbohydrate libraries (*see, e.g.,* Liang *et al., Science,* 274:1520-1522 (1996) and U.S. Patent 5,593,853), small organic molecule libraries (*see, e.g.,* benzodiazepines, Baum C&EN, Jan 18, page 33 (1993); isoprenoids, U.S. Patent 5,569,588; thiazolidinones and metathiazanones, U.S. Patent 5,549,974; pyrrolidines, U.S. Patents 5,525,735 and 5,519,134; morpholino compounds, U.S. Patent 5,506,337; benzodiazepines, 5,288,514, and the like).

Devices for the preparation of combinatorial libraries are commercially available (*see, e.g.,* 357 MPS, 390 MPS, Advanced Chem Tech, Louisville KY, Symphony, Rainin, Woburn, MA, 433A Applied Biosystems, Foster City, CA, 9050 Plus,

Millipore, Bedford, MA). In addition, numerous combinatorial libraries are themselves commercially available (*see, e.g.,* ComGenex, Princeton, N.J., Tripos, Inc., St. Louis, MO, 3D Pharmaceuticals, Exton, PA, Martek Biosciences, Columbia, MD, etc.).

C. Solid State and soluble high throughput assays

In one embodiment the invention provide soluble assays using molecules such as a domain such as ligand binding domain, an extracellular domain, a transmembrane domain (e.g., one comprising seven transmembrane regions and cytosolic loops), the transmembrane domain and a cytoplasmic domain, an active site, a subunit association region, etc.; a domain that is covalently linked to a heterologous protein to create a chimeric molecule; GPCR-B3; or a cell or tissue expressing GPCR-B3, either naturally occurring or recombinant. In another embodiment, the invention provides solid phase based *in vitro* assays in a high throughput format, where the domain, chimeric molecule, GPCR-B3, or cell or tissue expressing GPCR-B3 is attached to a solid phase substrate.

In the high throughput assays of the invention, it is possible to screen up to several thousand different modulators or ligands in a single day. In particular, each well of a microtiter plate can be used to run a separate assay against a selected potential modulator, or, if concentration or incubation time effects are to be observed, every 5-10 wells can test a single modulator. Thus, a single standard microtiter plate can assay about 100 (e.g., 96) modulators. If 1536 well plates are used, then a single plate can easily assay from about 100- about 1500 different compounds. It is possible to assay several different plates per day; assay screens for up to about 6,000-20,000 different compounds is possible using the integrated systems of the invention. More recently, microfluidic approaches to reagent manipulation have been developed, e.g., by Caliper Technologies (Palo Alto, CA).

The molecule of interest can be bound to the solid state component, directly or indirectly, via covalent or non covalent linkage e.g., via a tag. The tag can be any of a variety of components. In general, a molecule which binds the tag (a tag binder) is fixed to a solid support, and the tagged molecule of interest (e.g., the taste transduction molecule of interest) is attached to the solid support by interaction of the tag and the tag binder.

A number of tags and tag binders can be used, based upon known molecular interactions well described in the literature. For example, where a tag has a

natural binder, for example, biotin, protein A, or protein G, it can be used in conjunction with appropriate tag binders (avidin, streptavidin, neutravidin, the Fc region of an immunoglobulin, *etc.*) Antibodies to molecules with natural binders such as biotin are also widely available and appropriate tag binders; *see*, SIGMA Immunochemicals 1998 catalogue SIGMA, St. Louis MO).

Similarly, any haptenic or antigenic compound can be used in combination with an appropriate antibody to form a tag/tag binder pair. Thousands of specific antibodies are commercially available and many additional antibodies are described in the literature. For example, in one common configuration, the tag is a first antibody and the tag binder is a second antibody which recognizes the first antibody. In addition to antibody-antigen interactions, receptor-ligand interactions are also appropriate as tag and tag-binder pairs. For example, agonists and antagonists of cell membrane receptors (e.g., cell receptor-ligand interactions such as transferrin, c-kit, viral receptor ligands, cytokine receptors, chemokine receptors, interleukin receptors, immunoglobulin receptors and antibodies, the cadherein family, the integrin family, the selectin family, and the like; *see, e.g.*, Pigott & Power, *The Adhesion Molecule Facts Book I* (1993). Similarly, toxins and venoms, viral epitopes, hormones (e.g., opiates, steroids, etc.), intracellular receptors (e.g. which mediate the effects of various small ligands, including steroids, thyroid hormone, retinoids and vitamin D; peptides), drugs, lectins, sugars, nucleic acids (both linear and cyclic polymer configurations), oligosaccharides, proteins, phospholipids and antibodies can all interact with various cell receptors.

Synthetic polymers, such as polyurethanes, polyesters, polycarbonates, polyureas, polyamides, polyethyleneimines, polyarylene sulfides, polysiloxanes, polyimides, and polyacetates can also form an appropriate tag or tag binder. Many other tag/tag binder pairs are also useful in assay systems described herein, as would be apparent to one of skill upon review of this disclosure.

Common linkers such as peptides, polyethers, and the like can also serve as tags, and include polypeptide sequences, such as poly gly sequences of between about 5 and 200 amino acids. Such flexible linkers are known to persons of skill in the art. For example, poly(ethelyne glycol) linkers are available from Shearwater Polymers, Inc. Huntsville, Alabama. These linkers optionally have amide linkages, sulfhydryl linkages, or heterofunctional linkages.

Tag binders are fixed to solid substrates using any of a variety of methods currently available. Solid substrates are commonly derivatized or functionalized by

exposing all or a portion of the substrate to a chemical reagent which fixes a chemical group to the surface which is reactive with a portion of the tag binder. For example, groups which are suitable for attachment to a longer chain portion would include amines, hydroxyl, thiol, and carboxyl groups. Aminoalkylsilanes and hydroxyalkylsilanes can be used to functionalize a variety of surfaces, such as glass surfaces. The construction of such solid phase biopolymer arrays is well described in the literature. *See, e.g.*, Merrifield, *J. Am. Chem. Soc.* 85:2149-2154 (1963) (describing solid phase synthesis of, e.g., peptides); Geysen *et al., J. Immun. Meth.* 102:259-274 (1987) (describing synthesis of solid phase components on pins); Frank & Doring, *Tetrahedron* 44:60316040 (1988) (describing synthesis of various peptide sequences on cellulose disks); Fodor *et al., Science*, 251:767-777 (1991); Sheldon *et al., Clinical Chemistry* 39(4):718-719 (1993); and Kozal *et al., Nature Medicine* 2(7):753759 (1996) (all describing arrays of biopolymers fixed to solid substrates). Non-chemical approaches for fixing tag binders to substrates include other common methods, such as heat, cross-linking by UV radiation, and the like.

D. Computer-based assays

Yet another assay for compounds that modulate GPCR-B3 activity involves computer assisted drug design, in which a computer system is used to generate a three-dimensional structure of GPCR-B3 based on the structural information encoded by the amino acid sequence. The input amino acid sequence interacts directly and actively with a preestablished algorithm in a computer program to yield secondary, tertiary, and quaternary structural models of the protein. The models of the protein structure are then examined to identify regions of the structure that have the ability to bind, e.g., ligands. These regions are then used to identify ligands that bind to the protein.

The three-dimensional structural model of the protein is generated by entering protein amino acid sequences of at least 10 amino acid residues or corresponding nucleic acid sequences encoding a GPCR-B3 polypeptide into the computer system. The amino acid sequence of the polypeptide of the nucleic acid encoding the polypeptide is selected from the group consisting of SEQ ID NOS:1-3 or SEQ ID NOS:4-6 and conservatively modified versions thereof. The amino acid sequence represents the primary sequence or subsequence of the protein, which encodes the structural information of the protein. At least 10 residues of the amino acid sequence (or a nucleotide sequence encoding 10 amino acids) are entered into the computer system from computer

keyboards, computer readable substrates that include, but are not limited to, electronic storage media (e.g., magnetic diskettes, tapes, cartridges, and chips), optical media (e.g., CD ROM), information distributed by internet sites, and by RAM. The three-dimensional structural model of the protein is then generated by the interaction of the amino acid sequence and the computer system, using software known to those of skill in the art.

The amino acid sequence represents a primary structure that encodes the information necessary to form the secondary, tertiary and quaternary structure of the protein of interest. The software looks at certain parameters encoded by the primary sequence to generate the structural model. These parameters are referred to as "energy terms," and primarily include electrostatic potentials, hydrophobic potentials, solvent accessible surfaces, and hydrogen bonding. Secondary energy terms include van der Waals potentials. Biological molecules form the structures that minimize the energy terms in a cumulative fashion. The computer program is therefore using these terms encoded by the primary structure or amino acid sequence to create the secondary structural model.

The tertiary structure of the protein encoded by the secondary structure is then formed on the basis of the energy terms of the secondary structure. The user at this point can enter additional variables such as whether the protein is membrane bound or soluble, its location in the body, and its cellular location, e.g., cytoplasmic, surface, or nuclear. These variables along with the energy terms of the secondary structure are used to form the model of the tertiary structure. In modeling the tertiary structure, the computer program matches hydrophobic faces of secondary structure with like, and hydrophilic faces of secondary structure with like.

Once the structure has been generated, potential ligand binding regions are identified by the computer system. Three-dimensional structures for potential ligands are generated by entering amino acid or nucleotide sequences or chemical formulas of compounds, as described above. The three-dimensional structure of the potential ligand is then compared to that of the GPCR-B3 protein to identify ligands that bind to GPCR-B3. Binding affinity between the protein and ligands is determined using energy terms to determine which ligands have an enhanced probability of binding to the protein.

Computer systems are also used to screen for mutations, polymorphic variants, alleles and interspecies homologs of GPCR-B3 genes. Such mutations can be associated with disease states or genetic traits. As described above, GeneChip™ and

related technology can also be used to screen for mutations, polymorphic variants, alleles and interspecies homologs. Once the variants are identified, diagnostic assays can be used to identify patients having such mutated genes. Identification of the mutated GPCR-B3 genes involves receiving input of a first nucleic acid or amino acid sequence encoding GPCR-B3, selected from the group consisting of SEQ ID NOS:1-3, or SEQ ID NOS:4-6 and conservatively modified versions thereof. The sequence is entered into the computer system as described above. The first nucleic acid or amino acid sequence is then compared to a second nucleic acid or amino acid sequence that has substantial identity to the first sequence. The second sequence is entered into the computer system in the manner described above. Once the first and second sequences are compared, nucleotide or amino acid differences between the sequences are identified. Such sequences can represent allelic differences in GPCR-B3 genes, and mutations associated with disease states and genetic traits.

VIII. Kits

GPCR-B3 and its homologs are a useful tool for identifying taste receptor cells, for forensics and paternity determinations, and for examining taste transduction. GPCR-B3 specific reagents that specifically hybridize to GPCR-B3 nucleic acid, such as GPCR-B3 probes and primers, and GPCR-B3 specific reagents that specifically bind to the GPCR-B3 protein, e.g., GPCR-B3 antibodies are used to examine taste cell expression and taste transduction regulation.

Nucleic acid assays for the presence of GPCR-B3 DNA and RNA in a sample include numerous techniques are known to those skilled in the art, such as Southern analysis, northern analysis, dot blots, RNase protection, S1 analysis, amplification techniques such as PCR and LCR, and *in situ* hybridization. In *in situ* hybridization, for example, the target nucleic acid is liberated from its cellular surroundings in such as to be available for hybridization within the cell while preserving the cellular morphology for subsequent interpretation and analysis. The following articles provide an overview of the art of *in situ* hybridization: Singer *et al., Biotechniques* 4:230-250 (1986); Haase *et al., Methods in Virology*, vol. VII, pp. 189-226 (1984); and *Nucleic Acid Hybridization: A Practical Approach* (Hames *et al.*, eds. 1987). In addition, GPCR-B3 protein can be detected with the various immunoassay techniques described above. The test sample is typically compared to both a positive control (e.g., a sample expressing recombinant GPCR-B3) and a negative control.

The present invention also provides for kits for screening for modulators of GPCR-B3. Such kits can be prepared from readily available materials and reagents. For example, such kits can comprise any one or more of the following materials: GPCR-B3, reaction tubes, and instructions for testing GPCR-B3 activity. Optionally, the kit contains biologically active GPCR-B3. A wide variety of kits and components can be prepared according to the present invention, depending upon the intended user of the kit and the particular needs of the user.

IX. Administration and pharmaceutical compositions

Taste modulators can be administered directly to the mammalian subject for modulation of taste *in vivo*. Administration is by any of the routes normally used for introducing a modulator compound into ultimate contact with the tissue to be treated, e.g., the tongue or mouth. The taste modulators are administered in any suitable manner, optionally with pharmaceutically acceptable carriers. Suitable methods of administering such modulators are available and well known to those of skill in the art, and, although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route.

Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular method used to administer the composition. Accordingly, there is a wide variety of suitable formulations of pharmaceutical compositions of the present invention (*see, e.g., Remington's Pharmaceutical Sciences*, 17th ed. 1985)).

The taste modulators, alone or in combination with other suitable components, can be made into aerosol formulations (i.e., they can be "nebulized") to be administered via inhalation. Aerosol formulations can be placed into pressurized acceptable propellants, such as dichlorodifluoromethane, propane, nitrogen, and the like.

Formulations suitable for administration include aqueous and non-aqueous solutions, isotonic sterile solutions, which can contain antioxidants, buffers, bacteriostats, and solutes that render the formulation isotonic, and aqueous and non-aqueous sterile suspensions that can include suspending agents, solubilizers, thickening agents, stabilizers, and preservatives. In the practice of this invention, compositions can be administered, for example, by orally, topically, intravenously, intraperitoneally, intravesically or intrathecally. Optionally, the compositions are administered orally or nasally. The formulations of compounds can be presented in unit-dose or multi-dose

sealed containers, such as ampules and vials. Solutions and suspensions can be prepared from sterile powders, granules, and tablets of the kind previously described. The modulators can also be administered as part a of prepared food or drug.

The dose administered to a patient, in the context of the present invention should be sufficient to effect a beneficial response in the subject over time. The dose will be determined by the efficacy of the particular taste modulators employed and the condition of the subject, as well as the body weight or surface area of the area to be treated. The size of the dose also will be determined by the existence, nature, and extent of any adverse side-effects that accompany the administration of a particular compound or vector in a particular subject.

In determining the effective amount of the modulator to be administered in a physician may evaluate circulating plasma levels of the modulator, modulator toxicities,, and the production of anti-modulator antibodies. In general, the dose equivalent of a modulator is from about 1 ng/kg to 10 mg/kg for a typical subject.

For administration, taste modulators of the present invention can be administered at a rate determined by the LD-50 of the modulator, and the side-effects of the inhibitor at various concentrations, as applied to the mass and overall health of the subject. Administration can be accomplished via single or divided doses.

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to one of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

EXAMPLES

The following examples are provided by way of illustration only and not by way of limitation. Those of skill in the art will readily recognize a variety of noncritical parameters that could be changed or modified to yield essentially similar results.

Example I: Cloning and expression of GPCR-B3

Since taste transduction occurs in taste receptor cells found in taste buds of the tongue and palate epithelium, a full-length cDNA library was generated from rat taste papillae. This library was made by oligo-dT priming of poly-A+ RNA isolated from several hundreds rat circumvallate papillae using a directional lZAP vector (Stratagene Inc; Hoon & Ryba, *J. Dent. Res.* 76:831-838 (1997)) following standard molecular biology procedures (*see, e.g.*, Ausubel *et al., Current Protocols in Molecular Biology* (1995). A collection of single-cell and single taste-bud cDNA libraries was also generated from individually isolated taste receptor cells and taste buds from rat and mouse circumvallate, foliate and fungiform papillae according to the method of Dulac & Axel, *Cell* 83:195-206 (1995). Taste buds and single taste receptor cells were isolated by enzymatic digestion and micro-dissection of lingual epithelium from adult rats and mice. To maximize lysis efficiency in the taste bud preparations, the lysis buffer volume was increased 10 fold (Dulac & Axel, *supra*).

GPCR-B3 was isolated from the lZAP circumvallate cDNA library by first generating a subtracted library enriched in sequences expressed in taste tissue. Construction and initial analysis of a taste-receptor cell subtracted cDNA library was as described by Hoon & Ryba, *supra*. Further enrichment of taste-specific transcripts was achieved by dot-blot screening of cDNA clones with non-taste cDNA probes. Non-taste probes included lingual epithelium tissue devoid of taste buds, muscle, liver and brain tissue. The individual hybridization probes were generated by preparing first strand cDNA and labeling it using random priming methods as described in Ausubel *et al., supra*. Hybridization conditions and washes were 65°C, 2x SSC for hybridizations, and 65°C, 0.1x SSC for washing.

All cDNAs that showed taste tissue enrichment in the differential screens with taste and non-taste tissue were picked for DNA sequencing analysis using standard dideoxy-termination methods and an automated ABI sequencing machine. DNA sequences were subjected to data analysis using a variety of homology and structure prediction programs (e.g. blast at http://www.ncbi.nlm.nih.gov/ Tmpred at http://dot.imgen.bcm.tmc.edu:9331/seq-search/struc-predict.html). Individual cDNA clones that encoded novel sequences, sequences with some similarity to known signaling components, sequences with multiple predicted transmembrane domains, or sequences with known motifs such as SH2, SH3, PDZ, etc (see for example pfam at http://pfam.wustl.edu/) were chosen as candidates for further analysis.

Candidate cDNAs were assayed for taste-cell expression by *in situ* hybridization to tissue sections of rat tongue. Tissue was obtained from adult rats. Fresh frozen sections (14 mm) were attached to silanized slides and prepared for *in situ* hybridization as described by Ryba & Tirindelli, *Neuron* 19:371-379 (1997). All *in situ* hybridizations were carried out at high stringency (5x SSC, 50 % formamide, 72 °C). For single-label detection, signals were developed using alkaline-phosphatase conjugated antibodies to digoxigenin and standard chromogenic substrates (Boehringer Mannheim) as described by Ryba & Tirindelli, *supra*. Partial DNA sequencing reactions were performed on ~2000 subtracted and single-cell cDNA clones, and *in situ* hybridizations were carried out with ~400 different candidate cDNAs. This screen identified a number of genes expressed in taste receptor cells including a single clone encoding a 3' fragment of GPCR-B3.

Full-length rat GPCR-B3 was isolated from the lZAP rat circumvallate cDNA library following standard plaque hybridization protocols (Ausubel *et al., supra*). Approximately 2.5 x106 clones were plated at high density on LB plates (~100,000 phage/plate) and replica filters were hybridized with a radiolabeled GPCR-B3 probe at high stringency (65°C, 2x SSC). Positive clones were picked, retested, purified and characterized by DNA restriction mapping and sequencing analysis. Several full-length GPCR-B3 clones were isolated and characterized (see SEQ ID NOS:4-6 and the amino acid sequences that they encode, SEQ ID NOS:1-3).

The mouse interspecies homolog of rat GPCR-B3 was isolated by screening a mouse genomic Bac and l library (Genome Systems) at low and moderate stringency (48°C, 7x SSC and 55°C, 5x SSC). The clones were characterized by restriction mapping and DNA sequencing. A mouse cDNA was isolated by performing RACE reactions (Marathon Kit, Clonetech) using first-strand cDNA prepared from RNA isolated from mouse circumvallate and foliate papillae. The human homolog of GPCR-B3 was isolated from a human testis library (Clonetech Inc.) following the observation that other sensory receptors such as olfactory and visual receptors are expressed in testis (Axel & Dulac, *supra*). See Figure 1 for a topological map of GPCR-B3, showing the extracellular domain, seven transmembrane domains, and an intracellular or C-terminal domain.

Example II: Western blot and *in situ* analysis

To demonstrate specific expression of GPCR-B3 protein in taste cells, antibodies were generated against short peptides and GPCR-B3 fusion proteins. The peptides consisted of 18 amino acid residues from the N- or C-terminal end of the GPCR-B3 predicted protein (*see, e.g.*, SEQ ID NO:1 and 2). The fusion proteins consisted of GST-fusion polypeptides encompassing the entire N-terminal domain or the last 3 predicted transmembrane segments plus the C-term region. Fusions were generated using standard molecular techniques (Harlow & Lane, *Antibodies* (1988)). Peptides were fused to carrier proteins, immunized into rabbits, and the serum affinity purified and assayed as described by Cassill *et al., Proc. Nat'l Acad. Sci. USA* 88:11067-11070 (1991)).

Antibodies were tested for specificity by western-blot analysis of protein homogenates from circumvallate or fungiform papillae. The blots also contained liver and brain protein extracts as negative controls. For immunohistochemistry, frozen sections were prepared as described by Ryba & Tirindelli, *supra* for *in situ* hybridizations, except that blocking reactions used 10 % donkey immunoglobulin, 1 % bovine serum albumin, 0.3% Triton X-100. Sections were incubated in the appropriate dilution of anti-TR1 (1:100) for 12-18 hrs., and detected using fluorescein-conjugated donkey anti-rabbit secondary antibodies (Jackson Immunolaboratory). Taste buds were counter-stained with the F-actin marker BODIPYRTR-X phallacidin (Molecular Probes). As a control for these studies, anti-NCAM antibodies were also used. Fluorescent images were obtained using a Leica TSC confocal microscope with an argon-krypton laser. Pre-treatment of the antibodies with the peptide immunogen abolished staining. See Figures 2 and 3 for western blot and *in situ* analysis, respectively.

WHAT IS CLAIMED IS:

1. An isolated nucleic acid encoding a sensory transduction G-protein coupled receptor, the receptor comprising greater than about 70% amino acid identity to an amino acid sequence of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3.

2. The isolated nucleic acid of claim 1, wherein the nucleic acid encodes a receptor that specifically binds to polyclonal antibodies generated against SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3.

3. The isolated nucleic acid of claim 1, wherein the nucleic acid encodes a receptor that has G-coupled protein receptor activity.

4. The isolated nucleic acid of claim 1, wherein the nucleic acid encodes a receptor comprising an amino acid sequence of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3.

5. The isolated nucleic acid sequence of claim 1, wherein the nucleic acid comprises a nucleotide sequence of SEQ ID NO:4, SEQ ID NO:5, or SEQ ID NO:6.

6. The isolated nucleic acid of claim 1, wherein the nucleic acid is from a human, a mouse, or a rat.

7. The isolated nucleic acid of claim 1, wherein the nucleic acid is amplified by primers that selectively hybridize under stringent hybridization conditions to the same sequence as degenerate primer sets encoding amino acid sequences selected from the group consisting of:

IAWDWNGPKW (SEQ ID NO:7) and

LPENYNEAKC (SEQ ID NO:8).

8. The isolated nucleic acid of claim 1, wherein the nucleic acid encodes a receptor having a molecular weight of about between 92 kDa to about 102 kDa.

9. An isolated nucleic acid encoding a sensory transduction G-protein coupled receptor, wherein the nucleic acid specifically hybridizes under highly stringent conditions to a nucleic acid having the sequence of SEQ ID NO:4, SEQ ID NO:5, or SEQ ID NO:6.

10. An isolated nucleic acid encoding a sensory transduction G-protein coupled receptor, the receptor comprising greater than about 70% amino acid identity to a polypeptide having a sequence of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3, wherein the nucleic acid selectively hybridizes under moderately stringent hybridization conditions to a nucleotide sequence of SEQ ID NO:4, SEQ ID NO:5, or SEQ ID NO:6.

11. An isolated nucleic acid encoding an extracellular domain of a sensory transduction G-protein coupled receptor, the extracellular domain having greater than about 70% amino acid sequence identity to the extracellular domain of SEQ ID NO:1.

12. The isolated nucleic acid of claim 11, wherein the nucleic acid encodes the extracellular domain linked to a nucleic acid encoding a heterologous polypeptide, forming a chimeric polypeptide.

13. The isolated nucleic acid of claim 11, wherein the nucleic acid encodes the extracellular domain of SEQ ID NO:1.

14. An isolated nucleic acid encoding a transmembrane domain of a sensory transduction G-protein coupled receptor, the transmembrane domain comprising greater than about 70% amino acid sequence identity to the transmembrane domain of SEQ ID NO:1.

15. The isolated nucleic acid of claim 14, wherein the nucleic acid encodes the transmembrane domain linked to a nucleic acid encoding a heterologous polypeptide, forming a chimeric polypeptide.

16. The isolated nucleic acid of claim 14, wherein the nucleic acid encodes the transmembrane domain of SEQ ID NO:1.

17. The isolated nucleic acid of claim 14, wherein the nucleic acid further encodes a cytoplasmic domain comprising greater than about 70% amino acid identity to the cytoplasmic domain of SEQ ID NO:1.

18. The isolated nucleic acid of claim 17, wherein the nucleic acid encodes the cytoplasmic domain of SEQ ID NO:1.

19. An isolated sensory transduction G-protein coupled receptor, the receptor comprising greater than about 70% amino acid sequence identity to an amino acid sequence of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3.

20. The isolated receptor of claim 19, wherein the receptor specifically binds to polyclonal antibodies generated against SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3.

21. The isolated receptor of claim 19, wherein the receptor has G-protein coupled receptor activity.

22. The isolated receptor of claim 19, wherein the receptor has an amino acid sequence of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3.

23. The isolated receptor of claim 19, wherein the receptor is from a human, a rat, or a mouse.

24. An isolated polypeptide comprising an extracellular domain of a sensory transduction G-protein coupled receptor, the extracellular domain comprising greater than about 70% amino acid sequence identity to the extracellular domain of SEQ ID NO:1.

25. The isolated polypeptide of claim 24, wherein the polypeptide encodes the extracellular domain of SEQ ID NO:1.

26. The isolated polypeptide of claim 24, wherein the extracellular domain is covalently linked to a heterologous polypeptide, forming a chimeric polypeptide.

27. An isolated polypeptide comprising a transmembrane domain of a sensory transduction G-protein coupled receptor, the transmembrane domain comprising greater than about 70% amino acid sequence identity to the transmembrane domain of SEQ ID NO:1.

28. The isolated polypeptide of claim 27, wherein the polypeptide encodes the transmembrane domain of SEQ ID NO:1.

29. The isolated polypeptide of claim 27, further comprising a cytoplasmic domain comprising greater than about 70% amino acid identity to the cytoplasmic domain of SEQ ID NO:1.

30. The isolated polypeptide of claim 29, wherein the polypeptide encodes the cytoplasmic domain of SEQ ID NO:1.

31. The isolated polypeptide of claim 27, wherein the transmembrane domain is covalently linked to a heterologous polypeptide, forming a chimeric polypeptide.

32. The isolated polypeptide of claim 31, wherein the chimeric polypeptide has G-protein coupled receptor activity.

33. An antibody that selectively binds to the receptor of claim 19.

34. An expression vector comprising the nucleic acid of claim 1.

35. A host cell transfected with the vector of claim 34.

36. A method for identifying a compound that modulates sensory signaling in sensory cells, the method comprising the steps of:

(i) contacting the compound with a polypeptide comprising an extracellular domain of a sensory transduction G-protein coupled receptor, the extracellular domain comprising greater than about 70% amino acid sequence identity to the extracellular domain of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3; and

(ii) determining the functional effect of the compound upon the extracellular domain.

37. The method of claim 36, wherein the polypeptide is a sensory transduction G-protein coupled receptor, the receptor comprising greater than about 70% amino acid identity to a polypeptide encoding SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3.

38. The method of claim 37, wherein the polypeptide comprises an extracellular domain that is covalently linked to a heterologous polypeptide, forming a chimeric polypeptide.

39. The method of claim 37 or 38, wherein the polypeptide has G-protein coupled receptor activity.

40. The method of claim 36, wherein the extracellular domain is linked to a solid phase.

41. The method of claim 40, wherein the extracellular domain is covalently linked to a solid phase.

42. The method of claim 37 or 38, wherein functional effect is determined by measuring changes in intracellular cAMP, IP3, or Ca^{2+}.

43. The method of claim 36, wherein the functional effect is a chemical effect.

44. The method of claim 36, wherein the functional effect is a physical effect.

45. The method of claim 36, wherein the functional effected is determined by measuring binding of the compound to the extracellular domain.

46. The method of claim 36, wherein the polypeptide is recombinant.

47. The method of claim 36, wherein the polypeptide is from a rat, a mouse, or a human.

48. The method of claim 37, wherein the polypeptide comprises an amino acid sequence of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3.

49. The method of claim 37 or 38, wherein the polypeptide is expressed in a cell or cell membrane.

50. The method of claim 49, wherein the cell is a eukaryotic cell.

51. A method for identifying a compound that modulates sensory signaling in sensory cells, the method comprising the steps of:

(i) contacting the compound with a polypeptide comprising an extracellular domain of a sensory transduction G-protein coupled receptor, the transmembrane domain comprising greater than about 70% amino acid sequence identity to the extracellular domain of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3; and

(ii) determining the functional effect of the compound upon the transmembrane domain.

52. The method of claim 51, wherein the polypeptide comprises an transmembrane domain that is covalently linked to a heterologous polypeptide, forming a chimeric polypeptide.

53. The method of claim 52, wherein the chimeric polypeptide has G-protein coupled receptor activity.

54. The method of claim 51, wherein the functional effect is determined by measuring changes in intracellular cAMP, IP3, or Ca^{2+}.

55. The method of claim 51, wherein the functional effect is a chemical effect.

56. The method of claim 51, wherein the functional effect is a physical effect.

57. The method of claim 51, wherein the polypeptide is recombinant.

58. The method of claim 51, wherein the polypeptide is from a rat, a mouse, or a human.

59. The method of claim 51 or 52, wherein the polypeptide is expressed in a cell or cell membrane.

60. The method of claim 59, wherein the cell is a eukaryotic cell.

61. A method of making a sensory transduction G-protein coupled receptor, the method comprising the step of expressing the receptor from a recombinant expression vector comprising a nucleic acid encoding the receptor, wherein the amino acid sequence of the receptor comprises greater than about 70% amino acid identity to a polypeptide having a sequence of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3.

62. A method of making a recombinant cell comprising a sensory transduction G-protein coupled receptor, the method comprising the step of transducing the cell with an expression vector comprising a nucleic acid encoding the receptor, wherein the amino acid sequence of the receptor comprises greater than about 70% amino acid identity to a polypeptide having a sequence of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3.

63. A method of making an recombinant expression vector comprising a nucleic acid encoding a sensory transduction G-protein coupled receptor, the method comprising the step of ligating to an expression vector a nucleic acid encoding the receptor, wherein the amino acid sequence of the receptor comprises greater than about

70% amino acid identity to a polypeptide having a sequence of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3.

AMENDED CLAIMS

[received by the International Bureau on 11 November 1999 (11.11.99); original claim 51 amended; remaining claims unchanged (1 page)]

45. The method of claim 36, wherein the functional effected is determined by measuring binding of the compound to the extracellular domain.

46. The method of claim 36, wherein the polypeptide is recombinant.

47. The method of claim 36, wherein the polypeptide is from a rat, a mouse, or a human.

48. The method of claim 37, wherein the polypeptide comprises an amino acid sequence of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3.

49. The method of claim 37 or 38, wherein the polypeptide is expressed in a cell or cell membrane.

50. The method of claim 49, wherein the cell is a eukaryotic cell.

51. A method for identifying a compound that modulates sensory signaling in sensory cells, the method comprising the steps of:

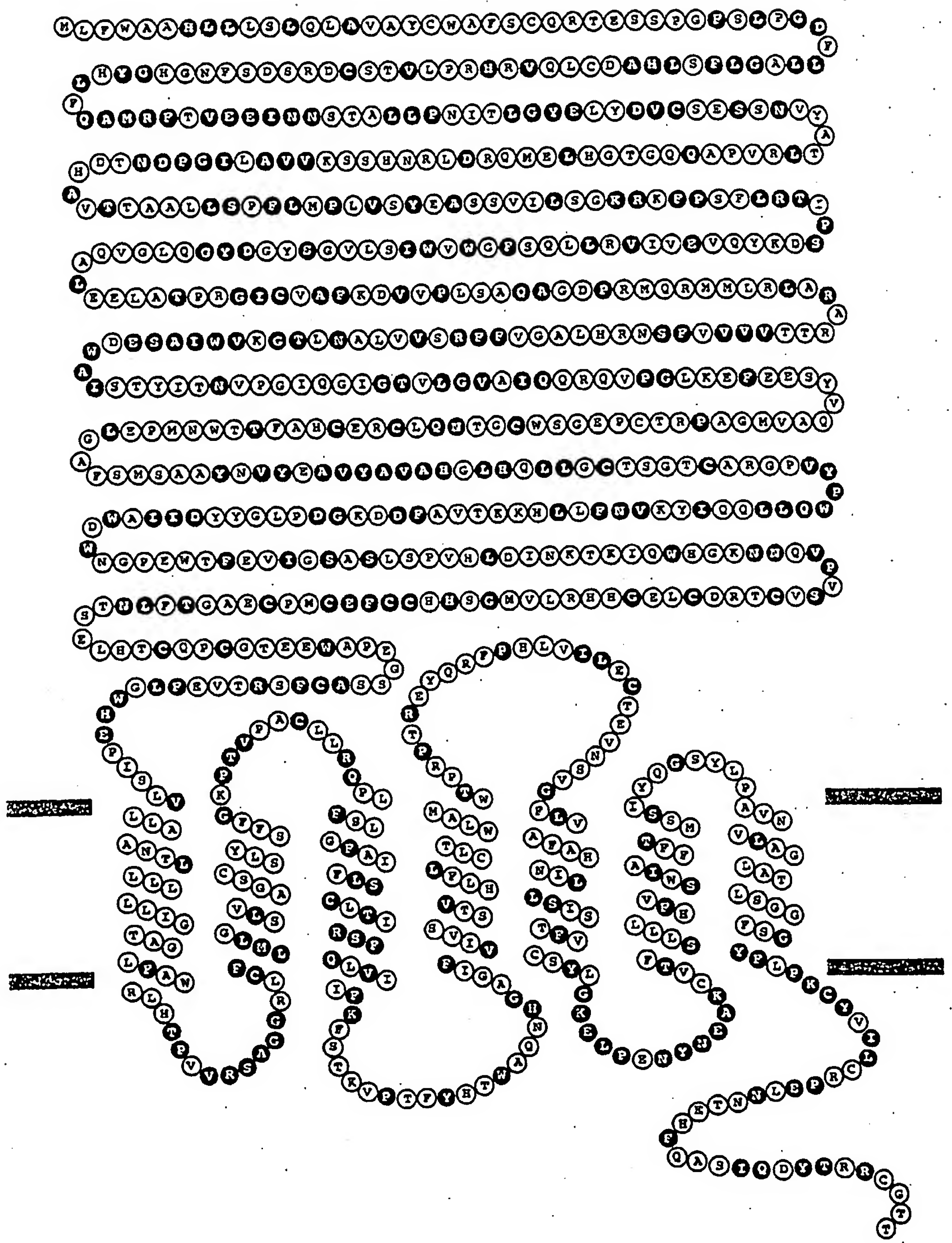
(i) contacting the compound with a polypeptide comprising a transmembrane domain of a sensory transduction G-protein coupled receptor, the transmembrane domain comprising greater than about 70% amino acid sequence identity to the transmembrane domain of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3; and

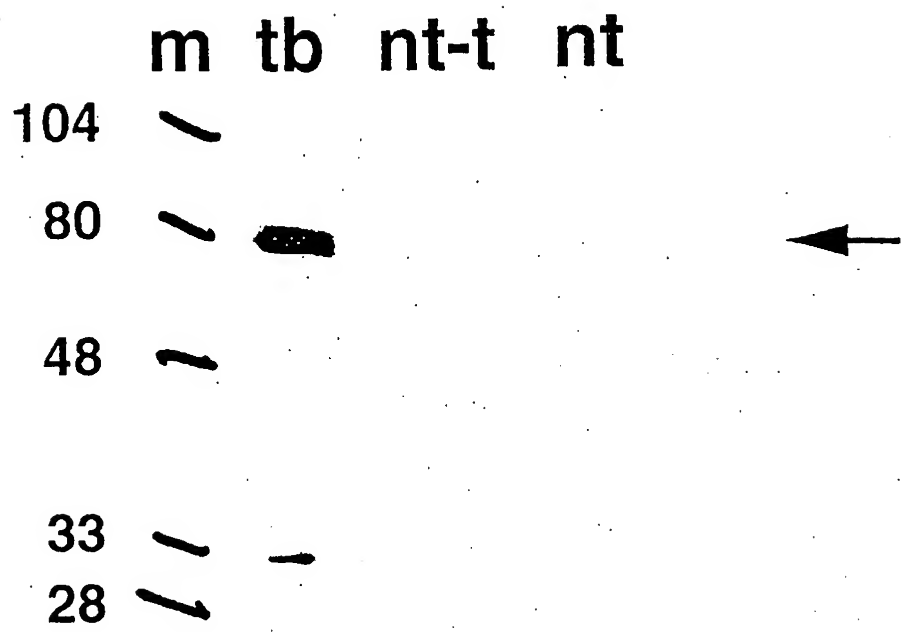
(ii) determining the functional effect of the compound upon the transmembrane domain.

52. The method of claim 51, wherein the polypeptide comprises an transmembrane domain that is covalently linked to a heterologous polypeptide, forming a chimeric polypeptide.

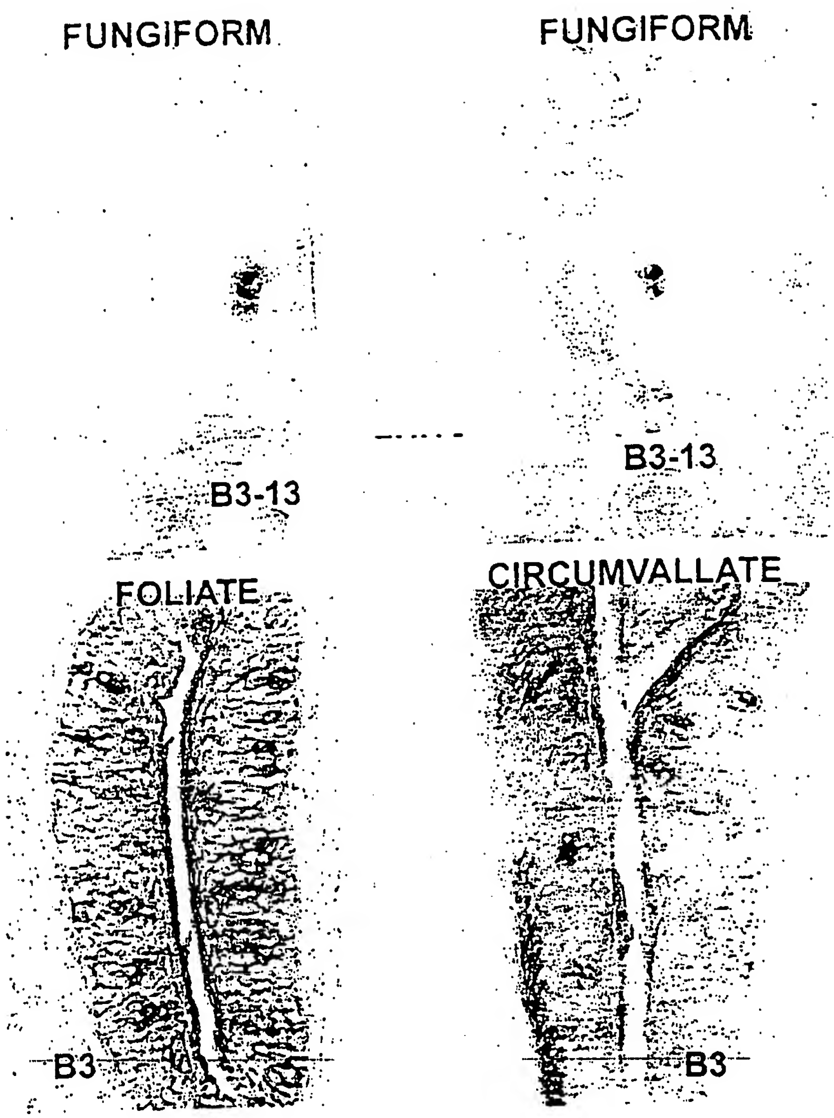
53. The method of claim 52, wherein the chimeric polypeptide has G-protein coupled receptor activity.

54. The method of claim 51, wherein the functional effect is determined by measuring changes in intracellular cAMP, IP3, or Ca^{2+}.

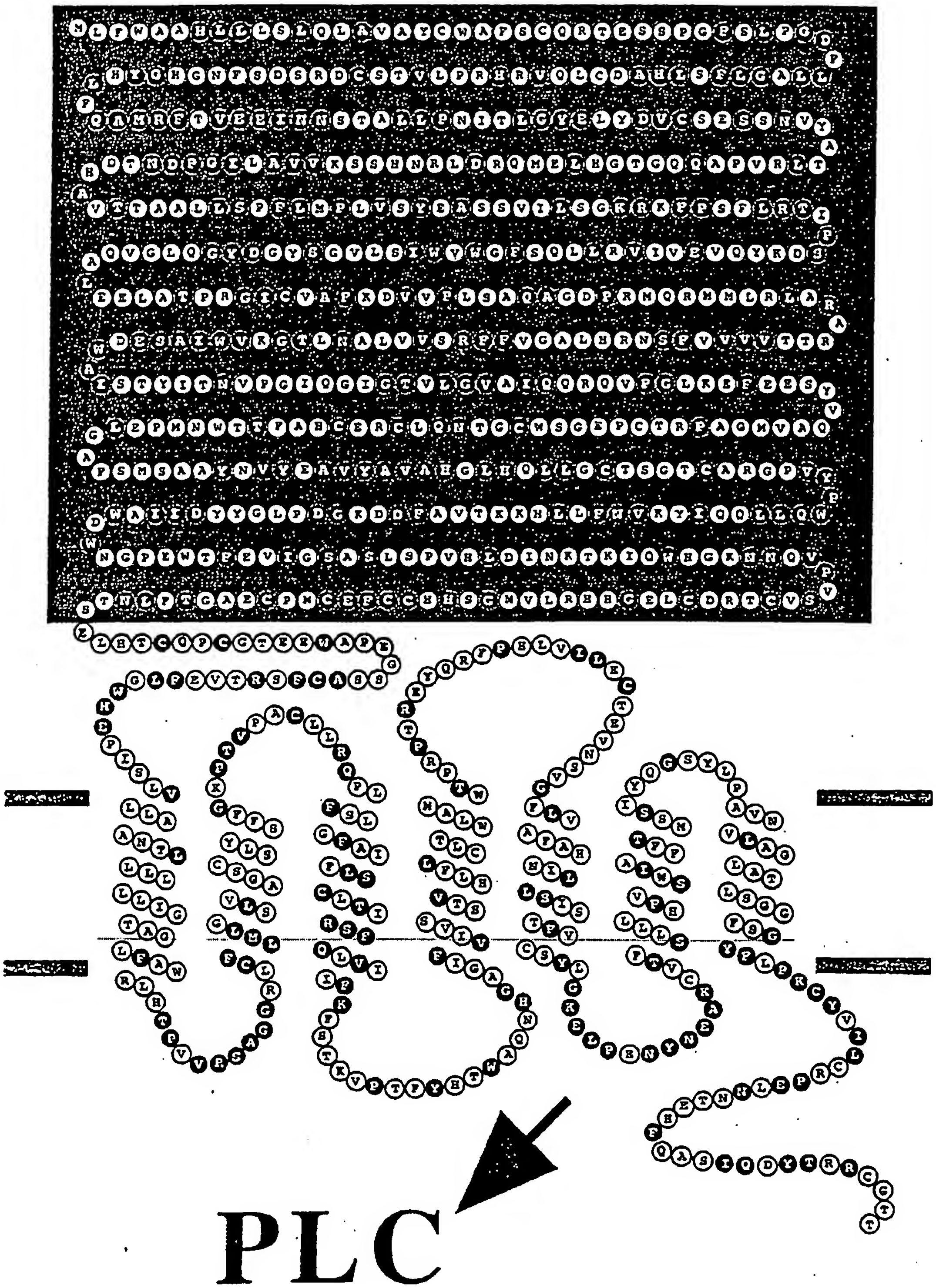


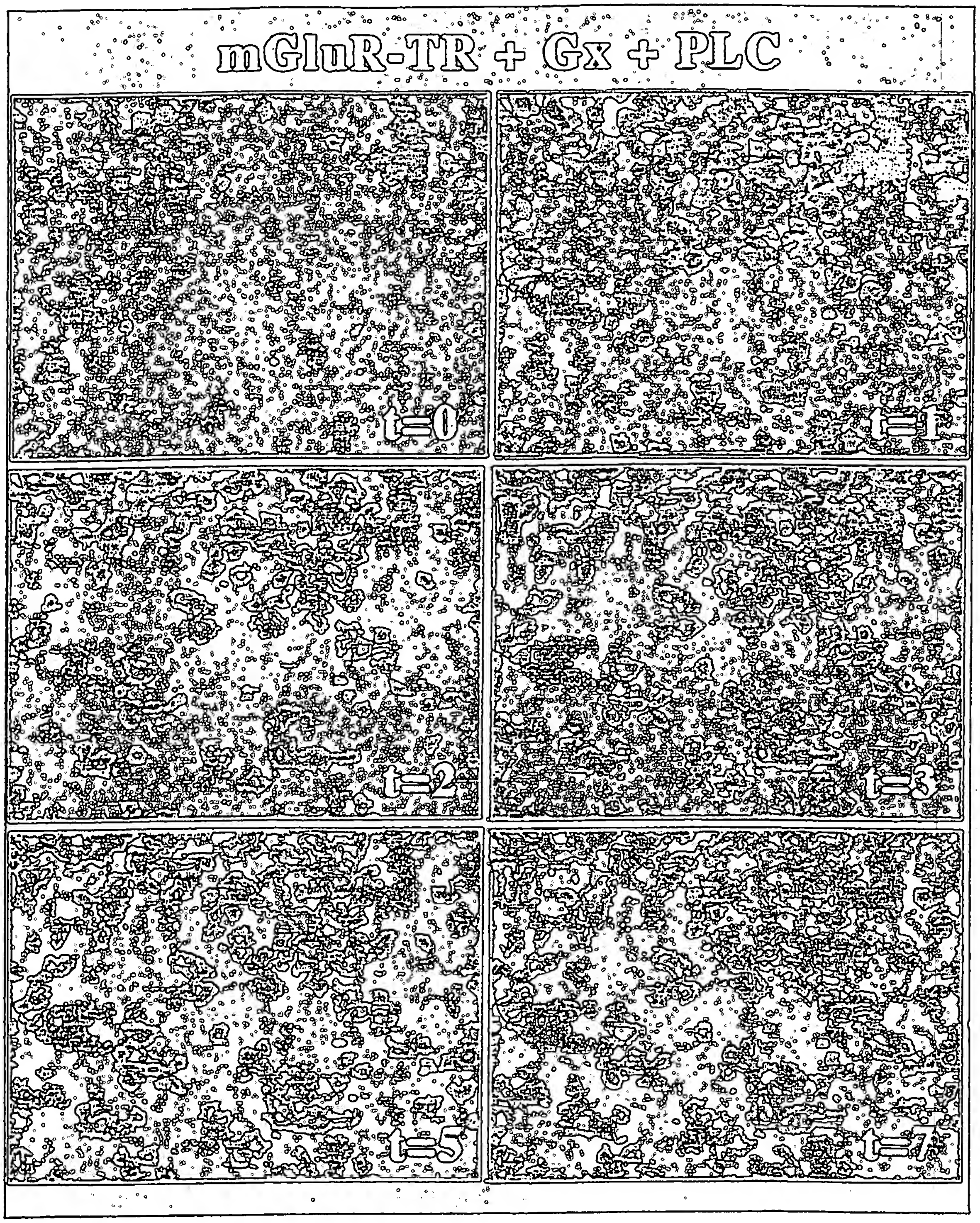


m=markers
tb=taste buds
nt-t=non-taste from tongue
nt=non-taste tissue


FUNGIFORM
FUNGIFORM
B3-13
B3-13
FOLIATE
CIRCUMVALLATE
B3
B3

mGluR-TR Chimeras




mGluR-TR + Gx + PLC
t=0
t=1
t=2
t=3
t=5
t=7

SEQUENCE LISTING

Rat GPCR-B3 amino acid sequence--SEQ ID NO:1

MLFWAAHLLLSLQLVYCWAFSCQRTESSPGFSLPGDFLLAGLFSLHGDCLQVRHRPLVTS
CDRPDSFNGHGYHLFQAMRFTVEEINNSSALLPNITLGYELYDVCSESANVYATLRVLAL
QGPRHIEIQKDLRNHSSKVVAFIGPDNTDHAVTTAALLGPFLMPLVSYEASSVVLSAKRK
FPSFLRTVPSDRHQVEVMVQLLQSFGWVWISLIGSYGDYGQLGVQALEELAVPRGICVAF
KDIVPFSARVGDPRMQSMMQHLAQARTTVVVVFSNRHLARVFFRSVVLANLTGKVWVASE
DWAISTYITSVTGIQGIGTVLGVAVQQRQVPGLKEFEESYVRAVTAAPSACPEGSWCSTN
QLCRECHTFTTRNMPTLGAFSMSAAYRVYEAVYAVAHGLHQLLGCTSEICSRGPVYPWQL
LQQIYKVNFLLHENTVAFDDNGDTLGYYDIIAWDWNGPEWTFEIIGSASLSPVHLDINKT
KIQWHGKNNQVPVSVCTTDCLAGHHRVVVGSHHCCFECVPCEAGTFLNMSELHICQPCGT
EEWAPKESTTCFPRTVEFLAWHEPISLVLIAANTLLLLLLVGTAGLFAWHFHTPVVRSAG
GRLCFLMLGSLVAGSCSFYSFFGEPTVPACLLRQPLFSLGFAIFLSCLTIRSFQLVIIFK
FSTKVPTFYRTWAQNHGAGLFVIVSSTVHLLICLTWLVMWTPRPTREYQRFPHLVILECT
EVNSVGFLLAFTHNILLSISTFVCSYLGKELPENYNEAKCVTFSLLLNFVSWIAFFTMAS
IYQGSYLPAVNVLAGLTTLSGGFSGYFLPKCYVILCRPELNNTEHFQASIQDYTRRCGTT

Mouse GPCR-B3 amino acid sequence--SEQ ID NO:2

MLFWAAHLLLSLQLAVAYCWAFSCQRTESSPGFSLPGDFLLAGLFSLHADCLQVRHRPLV
TSCDRSDSFNGHGYHLFQAMRFTVEEINNSTALLPNITLGYELYDVCSESSNVYATLRVP
AQQGTGHLEMQRDLRNHSSKVVALIGPDNTDHAVTTAALLSPFLMPLVSYEASSVILSGK
RKFPSFLRTIPSDKYQVEVIVRLLQSFGWVWISLVGSYGDYGQLGVQALEELATPRGICV
AFKDVVPLSAQAGDPRMQRMMLRLARARTTVVVVFSNRHLAGVFFRSVVLANLTGKVWIA
SEDWAISTYITNVPGIQGIGTVLGVAIQQRQVPGLKEFEESYVQAVMGAPRTCPEGSWCG
TNQLCRECHAFTTWNMPELGAFSMSAAYNVYEAVYAVAHGLHQLLGCTSGTCARGPVYPW
QLLQQIYKVNFLLHKKTVAFDDKGDPLGYYDIIAWDWNGPEWTFEVIGSASLSPVHLDIN
KTKIQWHGKNNQVPVSVCTRDCLEGHHRLVMGSHHCCFECMPCEAGTFLNTSELHTCQPC
GTEEWAPEGSSACFSRTVEFLGWHEPISLVLLAANTLLLLLLIGTAGLFAWRLHTPVVRS
AGGRLCFLMLGSLVAGSCSLYSFFGKPTVPACLLRQPLFSLGFAIFLSCLTIRSFQLVII
FKFSTKVPTFYHTWAQNHGAGIFVIVSSTVHLFLCLTWLAMWTPRPTREYQRFPHLVILE
CTEVNSVGFLVAFAHNILLSISTFVCSYLGKELPENYNEAKCVTFSLLLHFVSWIAFFTM
SSIYQGSYLPAVNVLAGLATLSGGFSGYFLPKCYVILCRPELNNTEHFQASIQDYTRRCG
TT

Human GPCR-B3 amino acid sequence--SEQ ID NO:3

RSCSFNEHGYHLFQAMRLGVEEINNSTALLPNITLGYQLYDVCSDSANVYATLRVLSLPG
QHHIELQGDLLHYSPTVLAVIGPDSTNRAATTAALLSPFLVHISYAASSETLSVKRQYPS
FLRTIPNDKYQVETMVLLLQKFGWTWISLVGSSDDYGQLGVQALENQALVRGICIAFKDI
MPFSAQVGDERMQCLMRHLAQAGATVVVVFSSRQLARVFFESVVLTNLTGKVWVASEAWA
LSRHITGVPGIQRIGMVLGVAIQKRAVPGLKAFEEAYARADKEAPRPCHKGSWCSSNQLC
RECQAFMAHTMPKLKAFSMSSAYNAYRAVYAVAHGLHQLLGCASELCSRGRVYPWQLLEQ
IHKVHFLLHKDTVAFNDNRDPLSSYNIIAWDWNGPKWTFTVLGSSTWSPVQLNINETKIQ
WHGKNHQVPKSVCSSDCLEGHQRVVTGFHHCCFECVPCGAGTFLNKSELYRCQPCGTEEW
APEGSQTCFPRTVVFLALREHTSWVLLAANTLLLLLLLGTAGLFAWHLDTPVVRSAGGRL
CFLMLGSLAAGSGSLYGFFGEPTRPACLLRQALFALGFTIFLSCLTVRSFQLIIIFKFST
KVPTFYHAWVQNHGAGLFVMISSAAQLLICLTWLVVWTPLPAREYQRFPHLVMLECTETN
SLGFILAFLYNGLLSISAFACSYLGKDLPENYNEAKCVTFSLLFNFVSWIAFFTTASVYD
GKYLPAANMMAGLSSLSSGFGGYFLPKCYVILCRPDLNSTEHFQASIQDYTRRCGST

Rat GPCR-B3 nucleotide sequence--SEQ ID NO:4

ATTCACATCAGAGCTGTGCTCAGCCATGCTGGGCAGAGGGACGACGGCTGGCCAGCATGC
TCTTCTGGGCTGCTCACCTGCTGCTCAGCCTGCAGTTGGTCTACTGCTGGGCTTTCAGCT
GCCAAAGGACAGAGTCCTCTCCAGGCTTCAGCCTTCCTGGGGACTTCCTCCTTGCAGGTC
TGTTCTCCCTCCATGGTGACTGTCTGCAGGTGAGACACAGACCTCTGGTGACAAGTTGTG
ACAGGCCCGACAGCTTCAACGGCCATGGCTACCACCTCTTCCAAGCCATGCGGTTCACTG
TTGAGGAGATAAACAACTCCTCGGCCCTGCTTCCCAACATCACCCTGGGGTATGAGCTGT
ACGACGTGTGCTCAGAATCTGCCAATGTGTATGCCACCCTGAGGGTGCTTGCCCTGCAAG
GGCCCCGCCACATAGAGATACAGAAAGACCTTCGCAACCACTCCTCCAAGGTGGTGGCCT
TCATCGGGCCTGACAACACTGACCACGCTGTCACTACCGCTGCCTTGCTGGGTCCTTTCC
TGATGCCCCTGGTCAGCTATGAGGCAAGCAGCGTGGTACTCAGTGCCAAGCGCAAGTTCC
CGTCTTTCCTTCGTACCGTCCCCAGTGACCGGCACCAGGTGGAGGTCATGGTGCAGCTGC
TGCAGAGTTTTGGGTGGGTGTGGATCTCGCTCATTGGCAGCTACGGTGATTACGGGCAGC
TGGGTGTGCAGGCGCTGGAGGAGCTGGCCGTGCCCCGGGGCATCTGCGTCGCCTTCAAGG
ACATCGTGCCTTTCTCTGCCCGGGTGGGTGACCCGAGGATGCAGAGCATGATGCAGCATC
TGGCTCAGGCCAGGACCACCGTGGTTGTGGTCTTCTCTAACCGGCACCTGGCTAGAGTGT
TCTTCAGGTCCGTGGTGCTGGCCAACCTGACTGGCAAAGTGTGGGTCGCCTCAGAAGACT
GGGCCATCTCCACGTACATCACCAGCGTGACTGGGATCCAAGGCATTGGGACGGTGCTCG

GTGTGGCCGTCCAGCAGAGACAAGTCCCTGGGCTGAAGGAGTTTGAGGAGTCTTATGTCA
GGGCTGTAACAGCTGCTCCCAGCGCTTGCCCGGAGGGGTCCTGGTGCAGCACTAACCAGC
TGTGCCGGGAGTGCCACACGTTCACGACTCGTAACATGCCCACGCTTGGAGCCTTCTCCA
TGAGTGCCGCCTACAGAGTGTATGAGGCTGTGTACGCTGTGGCCCACGGCCTCCACCAGC
TCCTGGGATGTACTTCTGAGATCTGTTCCAGAGGCCCAGTCTACCCCTGGCAGCTTCTTC
AGCAGATCTACAAGGTGAATTTTCTTCTACATGAGAATACTGTGGCATTTGATGACAACG
GGGACACTCTAGGTTACTACGACATCATCGCCTGGGACTGGAATGGACCTGAATGGACCT
TTGAGATCATTGGCTCTGCCTCACTGTCTCCAGTTCATCTGGACATAAATAAGACAAAAA
TCCAGTGGCACGGGAAGAACAATCAGGTGCCTGTGTCAGTGTGTACCACGGACTGTCTGG
CAGGGCACCACAGGGTGGTTGTGGGTTCCCACCACTGCTGCTTTGAGTGTGTGCCCTGCG
AAGCTGGGACCTTTCTCAACATGAGTGAGCTTCACATCTGCCAGCCTTGTGGAACAGAAG
AATGGGCACCCAAGGAGAGCACTACTTGCTTCCCACGCACGGTGGAGTTCTTGGCTTGGC
ATGAACCCATCTCTTTGGTGCTAATAGCAGCTAACACGCTATTGCTGCTGCTGCTGGTTG
GGACTGCTGGCCTGTTTGCCTGGCATTTTCACACACCTGTAGTGAGGTCAGCTGGGGGTA
GGCTGTGCTTCCTCATGCTGGGTTCCCTGGTGGCCGGAAGTTGCAGCTTCTATAGCTTCT
TCGGGGAGCCCACGGTGCCCGCGTGCTTGCTGCGTCAGCCCCTCTTTTCTCTCGGGTTTG
CCATCTTCCTCTCCTGCCTGACAATCCGCTCCTTCCAACTGGTCATCATCTTCAAGTTTT
CTACCAAGGTGCCCACATTCTACCGTACCTGGGCCCAAAACCATGGTGCAGGTCTATTCG
TCATTGTCAGCTCCACGGTCCATTTGCTCATCTGTCTCACATGGCTTGTAATGTGGACCC
CACGACCCACCAGGGAATACCAGCGCTTCCCCCATCTGGTGATTCTCGAGTGCACAGAGG
TCAACTCTGTAGGCTTCCTGTTGGCTTTCACCCACAACATTCTCCTCTCCATCAGTACCT
TCGTCTGCAGCTACCTGGGTAAGGAACTGCCAGAGAACTATAATGAAGCCAAATGTGTCA
CCTTCAGCCTGCTCCTCAACTTCGTATCCTGGATCGCCTTCTTCACCATGGCCAGCATTT
ACCAGGGCAGCTACCTGCCTGCGGTCAATGTGCTGGCAGGGCTGACCACACTGAGCGGCG
GCTTCAGCGGTTACTTCCTCCCCAAGTGCTATGTGATTCTCTGCCGTCCAGAACTCAACA
ATACAGAACACTTTCAGGCCTCCATCCAGGACTACACGAGGCGCTGCGGCACTACCTGAT
CCACTGGAAAGGTGCAGACGGGAAGGAAGCCTCTCTTCTTGTGCTGAAGGTGGCGGGTCC
AGTGGGGCCGAGAGCTTGAGGTGTCTGGGAGAGCTCCGGCACAGCTTACGATGTATAAGC
ACGCGGAAGAATCCAGTGCAATAAAGACGGGAAGTGTGAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAA

Mouse GPCR-B3 nucleotide sequence--SEQ ID NO:5

TTTGGCCAGCATGCTTTTCTGGGCAGCTCACCTGCTGCTCAGCCTGCAGCTGGCCGTTGC
TTACTGCTGGGCTTTCAGCTGCCAAAGGACAGAATCCTCTCCAGGTTTCAGCCTCCCTGG

GGACTTCCTCCTGGCAGGCCTGTTCTCCCTCCATGCTGACTGTCTGCAGGTGAGACACAG
ACCTCTGGTGACAAGTTGTGACAGGTCTGACAGCTTCAACGGCCATGGCTATCACCTCTT
CCAAGCCATGCGGTTCACCGTTGAGGAGATAAACAACTCCACAGCTCTGCTTCCCAACAT
CACCCTGGGGTATGAACTGTATGACGTGTGCTCAGAGTCTTCCAATGTCTATGCCACCCT
GAGGGTGCCCGCCCAGCAAGGGACAGGCCACCTAGAGATGCAGAGAGATCTTCGCAACCA
CTCCTCCAAGGTGGTGGCACTCATTGGGCCTGATAACACTGACCACGCTGTCACCACTGC
TGCCCTGCTGAGCCCTTTTCTGATGCCCCTGGTCAGCTATGAGGCGAGCAGCGTGATCCT
CAGTGGGAAGCGCAAGTTCCCGTCCTTCTTGCGCACCATCCCCAGCGATAAGTACCAGGT
GGAAGTCATAGTGCGGCTGCTGCAGAGCTTCGGCTGGGTCTGGATCTCGCTCGTTGGCAG
CTATGGTGACTACGGGCAGCTGGGCGTACAGGCGCTGGAGGAGCTGGCCACTCCACGGGG
CATCTGCGTCGCCTTCAAGGACGTGGTGCCTCTCTCCGCCCAGGCGGGTGACCCAAGGAT
GCAGCGCATGATGCTGCGTCTGGCTCGAGCCAGGACCACCGTGGTCGTGGTCTTCTCTAA
CCGGCACCTGGCTGGAGTGTTCTTCAGGTCTGTGGTGCTGGCCAACCTGACTGGCAAAGT
GTGGATCGCCTCCGAAGACTGGGCCATCTCCACGTACATCACCAATGTGCCCGGGATCCA
GGGCATTGGGACGGTGCTGGGGGTGGCCATCCAGCAGAGACAAGTCCCTGGCCTGAAGGA
GTTTGAAGAGTCCTATGTCCAGGCAGTGATGGGTGCTCCCAGAACTTGCCCAGAGGGGTC
CTGGTGCGGCACTAACCAGCTGTGCAGGGAGTGTCACGCTTTCACGACATGGAACATGCC
CGAGCTTGGAGCCTTCTCCATGAGCGCTGCCTACAATGTGTATGAGGCTGTGTATGCTGT
GGCCCACGGCCTCCACCAGCTCCTGGGATGTACCTCTGGGACCTGTGCCAGAGGCCCAGT
CTACCCCTGGCAGCTTCTTCAGCAGATCTACAAGGTGAATTTCCTTCTACATAAGAAGAC
TGTAGCATTCGATGACAAGGGGGACCCTCTAGGTTATTATGACATCATCGCCTGGGACTG
GAATGGACCTGAATGGACCTTTGAGGTCATTGGTTCTGCCTCACTGTCTCCAGTTCATCT
AGACATAAATAAGACAAAAATCCAGTGGCACGGGAAGAACAATCAGGTGCCTGTGTCAGT
GTGTACCAGGGACTGTCTCGAAGGGCACCACAGGTTGGTCATGGGTTCCCACCACTGCTG
CTTCGAGTGCATGCCCTGTGAAGCTGGGACATTTCTCAACACGAGTGAGCTTCACACCTG
CCAGCCTTGTGGAACAGAAGAATGGGCCCCTGAGGGGAGCTCAGCCTGCTTCTCACGCAC
CGTGGAGTTCTTGGGGTGGCATGAACCCATCTCTTTGGTGCTATTAGCAGCTAACACGCT
ATTGCTGCTGCTGCTGATTGGGACTGCTGGCCTGTTTGCCTGGCGTCTTCACACGCCTGT
TGTGAGGTCAGCTGGGGGTAGGCTGTGCTTCCTCATGCTGGGTTCCTTGGTAGCTGGGAG
TTGCAGCCTCTACAGCTTCTTCGGGAAGCCCACGGTGCCCGCGTGCTTGCTGCGTCAGCC
CCTCTTTTCTCTCGGGTTTGCCATTTTCCTCTCCTGTCTGACAATCCGCTCCTTCCAACT
GGTCATCATCTTCAAGTTTTCTACCAAGGTACCCACATTCTACCACACTTGGGCCCAAAA
CCATGGTGCCGGAATATTCGTCATTGTCAGCTCCACGGTCCATTTGTTCCTCTGTCTCAC
GTGGCTTGCAATGTGGACCCCACGGCCCACCAGGGAGTACCAGCGCTTCCCCCATCTGGT

GATTCTTGAGTGCACAGAGGTCAACTCTGTGGGCTTCCTGGTGGCTTTCGCACACAACAT
CCTCCTCTCCATCAGCACCTTTGTCTGCAGCTACCTGGGTAAGGAACTGCCGGAGAACTA
TAACGAAGCCAAATGTGTCACCTTCAGCCTGCTCCTCCACTTCGTATCCTGGATCGCTTT
CTTCACCATGTCCAGCATTTACCAGGGCAGCTACCTACCCGCGGTCAATGTGCTGGCAGG
GCTGGCCACTCTGAGTGGCGGCTTCAGCGGCTATTTCCTCCCTAAATGCTACGTGATTCT
CTGCCGTCCAGAACTCAACAACACAGAACACTTTCAGGCCTCCATCCAGGACTACACGAG
GCGCTGCGGCACTACCTGAGGCGCTGCGGCACTACCTGAGGCGCTGCGGCACTACCTGA

Human GPCR-B3 nucleotide sequence--SEQ ID NO:6

AGGTCTTGTAGCTTCAATGAGCATGGCTACCACCTCTTCCAGGCTATGCGGCTTGGGGTT
GAGGAGATAAACAACTCCACGGCCCTGCTGCCCAACATCACCCTGGGGTACCAGCTGTAT
GATGTGTGTTCTGACTCTGCCAATGTGTATGCCACGCTGAGAGTGCTCTCCCTGCCAGGG
CAACACCACATAGAGCTCCAAGGAGACCTTCTCCACTATTCCCCTACGGTGCTGGCAGTG
ATTGGGCCTGACAGCACCAACCGTGCTGCCACCACAGCCGCCCTGCTGAGCCCTTTCCTG
GTGCATATTAGCTATGCGGCCAGCAGCGAGACGCTCAGCGTGAAGCGGCAGTATCCCTCT
TTCCTGCGCACCATCCCCAATGACAAGTACCAGGTGGAGACCATGGTGCTGCTGCTGCAG
AAGTTCGGGTGGACCTGGATCTCTCTGGTTGGCAGCAGTGACGACTATGGGCAGCTAGGG
GTGCAGGCACTGGAGAACCAGGCCCTGGTCAGGGGCATCTGCATTGCTTTCAAGGACATC
ATGCCCTTCTCTGCCCAGGTGGGCGATGAGAGGATGCAGTGCCTCATGCGCCACCTGGCC
CAGGCCGGGGCCACCGTCGTGGTTGTTTTTCCAGCCGGCAGTTGGCCAGGGTGTTTTTC
GAGTCCGTGGTGCTGACCAACCTGACTGGCAAGGTGTGGGTCGCCTCAGAAGCCTGGGCC
CTCTCCAGGCACATCACTGGGGTGCCCGGGATCCAGCGCATTGGGATGGTGCTGGGCGTG
GCCATCCAGAAGAGGGCTGTCCCTGGCCTGAAGGCGTTTGAAGAAGCCTATGCCCGGGCA
GACAAGGAGGCCCCTAGGCCTTGCACAAGGGCTCCTGGTGCAGCAGCAATCAGCTCTGCA
GAGAATGCCAAGCTTTCATGGCACACACGATGCCCAAGCTCAAAGCCTTCTCCATGAGTT
CTGCCTACAACGCATACCGGGCTGTGTATGCGGTGGCCCATGGCCTCCACCAGCTCCTGG
GCTGTGCCTCTGAGCTCTGTTCCAGGGGCCGAGTCTACCCCTGGCAGCTTTTGGAGCAGA
TCCACAAGGTGCATTTCCTTCTACACAAGGACACTGTGGCGTTTAATGACAACAGAGATC
CCCTCAGTAGCTATAACATAATTGCCTGGGACTGGAATGGACCCAAGTGGACCTTCACGG
TCCTCGGTTCCTCCACATGGTCTCCAGTTCAGCTAAACATAAATGAGACCAAAATCCAGT
GGCACGGAAAGAACCACCAGGTGCCTAAGTCTGTGTGTTCCAGCGACTGTCTTGAAGGGC
ACCAGCGAGTGGTTACGGGTTTCCATCACTGCTGCTTTGAGTGTGTGCCCTGTGGGGCTG
GGACCTTCCTCAACAAGAGCGAGCTCTACAGATGCCAGCCTTGTGGAACAGAAGAGTGGG
CACCTGAGGGAAGCCAGACCTGCTTCCCGCGCACTGTGGTGTTTTTGGCTTTGCGTGAGC

ACACCTCTTGGGTGCTGCTGGCAGCTAACACGCTGCTGCTGCTGCTGCTGCTTGGGACTGCTGGCCTGTTTGCCTGGCACCTAGACACCCCTGTGGTGAGGTCAGCAGGGGGCCGCCTGTGCTTTCTTATGCTGGGCTCCCTGGCAGCAGGTAGTGGCAGCCTCTATGGCTTCTTTGGGGAACCCACAAGGCCTGCGTGCTTGCTACGCCAGGCCCTCTTTGCCCTTGGTTTCACCATCTTCCTGTCCTGCCTGACAGTTCGCTCATTCCAACTAATCATCATCTTCAAGTTTTCCACCAAGGTACCTACATTCTACCACGCCTGGGTCCAAAACCACGGTGCTGGCCTGTTTGTGATGATCAGCTCAGCGGCCCAGCTGCTTATCTGTCTAACTTGGCTGGTGGTGTGGACCCCACTGCCTGCTAGGGAATACCAGCGCTTCCCCCATCTGGTGATGCTTGAGTGCACAGAGACCAACTCCCTGGGCTTCATACTGGCCTTCCTCTACAATGGCCTCCTCTCCATCAGTGCCTTTGCCTGCAGCTACCTGGGTAAGGACTTGCCAGAGAACTACAACGAGGCCAAATGTGTCACCTTCAGCCTGCTCTTCAACTTCGTGTCCTGGATCGCCTTCTTCACCACGGCCAGCGTCTACGACGGCAAGTACCTGCCTGCGGCCAACATGATGGCTGGGCTGAGCAGCCTGAGCAGCGGCTTCGGTGGGTATTTTCTGCCTAAGTGCTACGTGATCCTCTGCCGCCCAGACCTCAACAGCACAGAGCACTTCCAGGCCTCCATTCAGGACTACACGAGGCGCTGCGGCTCCACCTGA

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/17099

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C07K 1/00; C07H 21/04; C12P 21/06
US CL : 530/350; 536/23.5; 435/691
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 530/350; 536/23.5; 435/691

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

STN: Biosis, Medline
Search terms: taste receptor, gustatory receptor, cDNA, clone, DNA

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 4,146,501 A (HENKIN) 27 March 1979.	1-32, 34, 35, 61-63
A	US 5,688,662 A (MARGOLSKEE) 18 November 1997.	1-32, 34, 35, 61-63
A	MARGOLSKEE, R.F. The molecular biology of taste transduction. BioEssays. October 1993, Vol. 15, No. 10, pages 645-650.	1-32, 34, 35, 61-63
A	ABE, K. et al. Primary structure and cell-type specific expression of a gustatory G protein-coupled receptor related to olfactory receptors. J. Biol. Chem. 05 June 1993, Vol. 268, No. 16, pages 12033-12039.	1-32, 34, 35, 61-63

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
08 OCTOBER 1999

Date of mailing of the international search report
04 NOV 1999

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer
MICHAEL T. BRANNOCK
Telephone No. (703) 308-0196

JOYCE BRIDGERS
PARALEGAL SPECIALIST
CHEMICAL MATRIX

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/17099

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-32, 34, 35, 61-63

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claim(s)1-32, 34, 35, 61-63, drawn to nucleic acids, vectors, host cells, polypeptides, and methods for making host cells and polypeptides.
Group II, claim(s) 33, drawn to an antibody.
Group III, claim(s) 36-60, drawn to methods of identifying modulators of sensory signal transduction.